ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/330113204

Mind and Formal Structures. The Kripke-Ross Argument Against Naturalizing Formal Understanding (Doctoral Dissertation, University of Leuven)

Thesis · January 2019

DOI: 10.16140/RG.5.6.11107.14883

CITATION
1

READS
1,171

1 author:

Antonio Ramos Diaz
Pontifical Catholic University of Puerto Rico
5 PUBLICATIONS **2** CITATIONS

SEE PROFILE

All content following this page was uploaded by Antonio Ramos Diaz on 02 July 2019.

The user has requested enhancement of the downloaded file.

KU Leuven
Groep Humane Wetenschappen
Hoger Instituut voor Wijsbegeerte

MIND AND FORMAL STRUCTURES:

ON THE KRIPKE-ROSS ARGUMENT AGAINST NATURALIZING FORMAL UNDERSTANDING

Antonio RAMOS DIAZ

Promotor:
Prof. Henning Tegtmeyer

Co-Promotors:
Prof. Howard Robinson
Prof. Jeff Buechner

Proefschrift voorgedragen tot het behalen van de graad van Doctor in de Wijsbegeerte

November 2018

KU Leuven
Humanities and Social Sciences Group
Institute of Philosophy

MIND AND FORMAL STRUCTURES:

ON THE KRIPKE-ROSS ARGUMENT AGAINST NATURALIZING FORMAL UNDERSTANDING

Antonio RAMOS DIAZ

Supervisor:
Prof. Henning Tegtmeyer

Co-Supervisors:
Prof. Howard Robinson
Prof. Jeff Buechner

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Philosophy

November 2018

To my parents, Sandra and Victor, the heroes of the epic drama of my life, for their unconditional love and support, and to my niece, Yamini, for having taught me that I was utterly oblivious to what beauty really was until the moment I held her in my arms.

In memory of my grandfather, Gilberto, a man for all ages.

Note to the present version

This version of my doctoral dissertation consists of minor revisions that were carried out during a period of six weeks, from the moment it was officially submitted up to three weeks before the doctoral defense. The revision consists almost entirely of corrections to grammatical and formatting errors (e.g., typos, poorly written and worded sentences, mistaken and missing references, etc.). I have, however, substantially augmented some of the sections, adding further references, caveats, and explanations in places where it seemed that sufficient attention and sensitivity to the literature and texts being discussed was lacking. I have not revised everything. Only some select parts of the dissertation have been revisited. The revision has been undertaken in such a way so as not to change the substance of what was in the original version submitted in partial fulfillment for the degree of Doctor in Philosophy. In a span of six weeks I have changed my mind about some of the arguments developed in the dissertation, yet I have opted to leave them untouched. This is the last phase of editing to which this dissertation will be subject.

Antonio Ramos–Díaz
KU Leuven, Leuven
January 1, 2019

TABLE OF CONTENTS

ACKNOWLEDGMENTS 1

INTRODUCTION I

1. THE CONCEPTUAL FRAMEWORK 1

1. THREE MAIN CLAIMS 1
2. THE CONSTITUTIVE AND THE GENETIC PROBLEM 1
3. THE UPSHOT OF THE ARGUMENT 2
4. SETTING THE STAGE: THE CONCEPTUAL FRAMEWORK 4
5. WHY FORMAL UNDERSTANDING? 29
6. PRACTICE-BOUND UDERSTANDING: PRACTICES, TRAINING, RULES 34
7. THE ARGUMENT: KRA 40

2. KRIPKE AND THE FORMAL INDETERMINACY OF THE PHYSICAL 41

1. THE ORIGINS: RULE-FOLLOWING AND THE SCEPTICAL PARADOX 42
2. WHAT IS COMPUTATION? RULE-FOLLOWING CONSIDERATIONS, FUNCTIONALISM, COMPUTATIONALISM, AND COMPUTATIONAL FUNCTIONALISM 52
3. THE ARGUMENT FROM DETERMINATE REALIZABILITY 65
4. KRIPKE AND PHYSICAL COMPUTING MECHANISMS 70
5. PHYSICAL COMPUTING MECHANISMS AND THE INEVITABILITY OF FORMAL INDETERMINACY 76

3. THE KRIPKEAN CHALLENGE: OBJECTIONS AND REPLIE 85

1. THE CONJUNCTIVE AND DISJUNCTIVE OBJECTIONS 85
2. LANDE ON UNDERDETERMINATION VS. INDETERMINACY 93
3. STABLER'S WAY OUT: THE CAUSAL-COUNTERFACTUAL ACCOUNT 96
4. WHAT ABOUT SEMANTICS? SEMANTICS AND COMPUTATION 105
5. THE SUI GENERIS PHYSICAL FACT OBJECTION 112
6. THE SLIPPERY SLOPE OBJECTION 113
7. IS THERE PHYSICAL COMPUTATION AT ALL? 115
8. CONCLUSION: THE FALLACY OF SOPHOMORIC RELATIVISM 120

4. ROSS AND THE FORMAL DETERMINACY OF UNDERSTANDING 125

1. KRIPKE'S ARGUMENT AND FOR FORMAL UNDERSTANDING 126
2. AN INTUITIVE SKETCH: THE FORMAL DETERMINACY OF UNDERSTANDING 128

3. CONTRA DENNETT: FORMAL UNDERSTANDING AND THE INTENTIONAL STANCE 138
4. RETREAT FROM PEOPLE 142
5. THE PARALLEL ARGUMENT AS TO CONTENT 151
6. MISUNDERSTANDING ROSS: THE FESER-DILLARD DEBATE 156
7. CONCLUSION 160

5. OBJECTIONS AND REPLIES 161

1. THE EPISTEMIC OBJECTION 161
2. THE OBJECTION FROM NON-STANDARD MODELS: SKOLEM AND KRA 165
3. THE PSYCHOLOGISM OBJECTION 167
4. THE ABSENT POSITIVE FACT OBJECTION 168
5. THE MEREOLOGICAL OBJECTION 169
6. THE CONTENT FALLACY OBJECTION 170
7. PUTNAM'S OBJECTION 174
8. DILLARD'S OBJECTION: MITOSIS OR SCHMITOSIS? 180
9. CONCLUSION: KRIPKE, ROSS, AND RULE-FOLLOWING 187

6. KRA, LANGUAGE, AND LINGUISTIC UNDERSTANDING 191

1. CHOMSKY ON THE NATURE OF LANGUAGE, THE LANGUAGE FACULTY, AND LINGUISTIC UNDERSTANDING 195
2. KRA AND CHOMSKY'S CONJUNCTION 206
3. CHOMSKY ON RULES AND RULE-FOLLOWING 212
4. CONCLUDING REMARKS 220

CONCLUSION 224

BIBLIOGRAPHY 229

ACKNOWLEDGMENTS

I would like to start by thanking Prof. Henning Tegtmeyer, for his willingness and diligence in supervising my thesis, for giving me the freedom to pursue my philosophical inklings, and for the unwavering support and encouragement he showed me throughout my doctoral research. I have benefited greatly from his guidance and feedback. I am immensely grateful to him. Special thanks to Prof. Howard Robinson and to Prof. Jeff Buechner, my co-supervisors, for their willingness to take on the task of co-supervising my research at a distance, for their helpful comments and criticisms without which this thesis would have been way worse than what it already is, and for their much needed words of encouragement and support. I thank you. Thanks to Prof. Buechner for having made one of Kripke's unpublished manuscripts available to me for this monograph, and special thanks to Saul Kripke for giving me permission to quote from that manuscript.

My ode of academic gratitude would not be complete if I do not mention three professors from the University of Puerto Rico (UPR) who together played the most fundamental role in my intellectual vocation: Prof. Guillermo Rosado Haddock (Philosophy), Prof. Mayra Rosario Urrutia (History), and Prof. Luis Ortiz (Linguistics). To Prof. Guillermo Rosado I owe my passion for logic and philosophy of mathematics, and all things related to logic and mathematics. He enthusiastically encouraged me to pursue graduate studies both in St. Andrews and KU Leuven, and made a 20-year-old kid from Carolina, Puerto Rico believe that he had what it took to be in the "big leagues". He remains, 'til this day, one of the best philosophy professors I have ever had. Words cannot express my gratitude. I also need to thank Prof. Mayra Rosario Urrutia, whose support and words of encouragement remain unparalleled, who was willing to write me more than 10+ letters of recommendation during the many applications I made for graduate schools as I was finishing my BA. After having submitted my BA paper (called *tesina*) in History and expressing doubts about an academic career, she said: "What? You are brilliant! Continue!". So I did. Not sure I would've continued had she not said what she said. Words cannot express how thankful I am. Last but not least is Prof. Luis Ortiz, who introduced me to the fascinating field of linguistics and always encouraged me to pursue it. Although I may have disappointed him when I opted for philosophy, I never forgot about linguistics. The present doctoral dissertation, with its recurring interest in linguistics and the nature of language, could not have turned out as it did had it not been for the impression his undergrad course in linguistics made on me. For his immense support, encouragement, and advice during one of the most pivotal periods of my life, I am forever grateful.

Last but not least, my family. There is no such thing as an intellectual in a vacuum. This all the more true if the intellectual is Puerto Rican! Were it not for my family, I don't think I would be writing these very words today. First and foremost, I need to thank the heroes of the epic drama that has been my life: my mother Sandra Díaz Menendez and my father Victor Antonio Ramos Rodriguez. No epic drama could have had better heroes and no son could have asked for better parents. It was no easy task to raise me. Yet somehow they excelled, and here I am today, not a high school or a college dropout (which is the outcome most people expected), but a soon to be doctor. The story of this roller-coaster of a life and the complete turnaround is better left for another occasion. Here I simply wish to point out that I would not have made it this far had it not been for them. Papi y mami, this dissertation is as much mine as it is yours. You are the GOAT, the *real* MVPs. To them, I dedicate this dissertation. ¡*Los amo*! To my brother Victor Ramos Díaz for being the best brother anyone could ever have. I love you, and though I have never told you this, I admire you. I pray and hope that one day I can be as good a man, son, brother, and father as you. To my late grandfather Gilberto Díaz Quiñones, a man of few words but of great spirit, presence, and courage. He was the only grandfather figure I had growing up and I cannot remember a single

important moment in my life when he wasn't present. For his love, for always being there, for the discipline he instilled in me and my brother, I thank him deeply. He passed away before he could see me complete and finish my doctoral studies. Having to pay him my last goodbyes *via* Facetime was no easy thing. Much has been sacrificed in the process. I hope to have made him proud every step of the way. Abuelo, I love you and I miss you. To my grandmothers, Alicia Menéndez Narváez and Edith Rodriguez Rosario, the cornerstones of the family and the strongest and most beautiful grandmas in the world, of course; I love you dearly. To my whole family, I love you all. Finally, to beauty itself made flesh, my niece Yamini Ramos Bergollo. You are too young to understand, but one day you will. Every page of this dissertation was written with a picture of you in front of me at all times. To her, the most beautiful being I have ever seen, I also dedicate this dissertation.

Antonio Ramos Díaz
KU Leuven, Leuven
November 19, 2018

INTRODUCTION

Contemporary philosophy of mind is a fascinating but perplexing field of study. It deals with what should be more known to us: our minds and cognitive capacities. Yet what should be more known to us turns out to be one of the most mindboggling and perplexing phenomena. Jerry Fodor (Fodor, 1996; see also Nagel, 2012) aptly described the philosophy of mind as a conjunction of three distinct problems:

1. the problem of *consciousness* ('qualia', our conscious, phenomenal, qualitative feel and experiences),
2. the problem of *intentionality* (the aboutness or directedness of certain mental states like desires) and
3. the problem of *rationality* (e.g., rational action, concepts and the logical relations that obtain amongst them, language, etc.).[1]

The problem of rationality can be split into two different sub-problems: one having to do with rational action, that is, acting on account of reasons; the other having to do with the nature of thought and understanding, particularly higher-order or abstract understanding, which carry semantic features like conceptual meaning and truth-values, and formal features like validity and mathematical form. The present work falls right on the latter sub-problem of the problem of rationality. We shall call this sub-problem 'the problem of explaining the nature of human *formal understanding*' (the formal aspects of the cogitative powers of human beings). The presumption shared by most contemporary philosophers is that problems 1, 2, and 3 must have a naturalistic, i.e. materialist, solution. "…these days we're all materialists for much the reason that Churchill gave for being a democrat: the alternatives seem even worse. Correspondingly, there's a new research agenda: to reconcile our materialism with the psychological facts; to explain how minds qua material objects could have the properties they do." (Fodor, 1996, p. 292)

Traditional debates over the overall cogency of materialism or physicalism as an intelligible and plausible framework for explaining the mind have for the most part focused on 1 and 2, and primarily on 1, the so-called "hard problem" of consciousness (see e.g. Chalmers, 1995a, 1997, 2010; Jackson, 1998; Ludlow et al., 2004; Strawson, 2008; Harnad, 2012a; for useful anthologies and textbooks on contemporary philosophy of mind, see Moser and Trout, 1995; Heil, 2004; Jackson and Smith, 2007; Nida-Rümelin, 2015; Beakley and Ludlow, 2006; McLaughlin et al., 2009; for some arguments against materialism about 1 and 2, see Göcke, 2012; Robinson, 1996; Plantinga, 2011). Some influential philosophers and cognitive scientists (e.g., Chalmers, 1995a; Levine, 2001; Kim, 2011; Harnad, 2012b) have come to swept 3 under the rug of the "easy problem" of consciousness, a problem in principle amenable to a materialist solution via a functionalist-, or strictly computationalist-, framework, although this claim has not gone unchallenged (see Kurzweil, 2012 for a popular example of the mind as a computer; for some famous arguments against computationalism that have very little in common, see e.g. Penrose, 1996a, 1999, 2002; Dreyfus, 1979, 1992; Searle, 1980, 1990, 1992). But these challenges are generally countenanced under the presumption that some other framework for naturalism, for naturalizing the mind, is possible. The

[1] This is not to suggest that the problem of rationality is separate or separable from the problem of intentionality, particularly when it comes to the nature of concepts and representations as Fodor himself has repeatedly made clear (Fodor, 1996, cf. 1981, 1983, 1985, 1994, 1998). Fodor is also known for having championed the semantic view of computation and for having coined the now famous expression "no computation without representation"(Fodor, 1981, p. 181).

received view is that rationality, in comparison to the subjective character of conscious experience, poses no special threat to naturalism (Chalmers 1995, 1996a; Levine 2001). Functionalism or computationalism (or a mix of both) - so goes the received view - has provided us the conceptual framework with which to begin to explain rationality in a way compatible with physicalism and the methods of the cognitive sciences (Chalmers 1996b, 2010; Fodor 1996; cf. Block 1980). This is because, according to the received view, all that an explanation of rationality involves is the explanation of cognitive functions, and in order to explain a cognitive function all one needs to do is specify a (e.g. physical) mechanism that can perform that function (Chalmers, 1995, p. 202). There are many different versions of functionalism (Levin, 2018; see e.g. Van Gulick, 2009) many of which are non-computational. In general, as a metaphysical theory of the mind, a functionalist theory will identify mental states with functional states, where a functional state is a state defined abstractly in terms of what it does in a given system rather than by what it is made of—since many disparate physical structures can realize or perform it (Levin 2018). Thus, mental states are not identical with any type of physical structure but rather with abstract functional structures that can be realized or performed by disparate physical mechanisms. A further widely accepted claim by functionalists is that such abstract functions are realized or performed *only* by physical structures (cf. Rey, 1997) such that all there is to having a mind (or being in a mental state) is to *physically* realize or perform the appropriate abstract functions and functional role. So functionalism, though itself ontologically neutral, generally comes hand-in-hand with token physicalism (Block 1980a, van Gulick 2009, Levin 2018), what I call 'the naturalistic commitment' of functional and computational theories of the mind. When it comes to the explanation of that aspect of our rationality that is most abstract (our capacity for formal, i.e. mathematical and logical, understanding) the standard functionalist or naturalist framework tends to be, by default, *computational* in which case mathematical and logical thinking can be accounted for in terms of physically realized computations, often in analogy with a (stochastic) Turing machine or concrete digital computers (Piccinini, 2009, 2010b, 2010a). The origins of this computational functionalist theory go back to the early 60s (Putnam, 1960, 1967a, 1967b; Fodor, 1965, 1968a, 1975, 1983; cf. Horst and Zalta, 2003; Rescorla, 2017).

The present work consists in a critical examination of the claim that human formal understanding (understanding of formal concepts, rules, and operations) and activity (applying formal rules, carrying out formal operations) is in some way wholly material and therefore in principle explicable in terms of purely physical or material capacities, abilities, and activities—regardless of whether the account of physical determination of formal properties is reductive or non-reductive. Our argument will attempt to refute non-reductive accounts of formal understanding, particularly functionalist and computationalist accounts as these provide the best naturalistic framework for explaining the *formal* potentialities and features of the cogitative capacities, abilities, and activities in question (i.e., our capacity and abilities for logic and mathematics). If the claim that these capacities, abilities, and activities are not wholly a matter of matter is correct, then we take our argument to show that there cannot be such a thing as a naturalistic explanation of human formal understanding in the sense in which there can be naturalistic explanations of purely physical, material phenomena (e.g., the physical, the chemical, the biological, etc.). In other words, even if we idealize to the final and explanatorily complete state of the natural sciences, no conjunction of complete or finalized natural sciences would contain a complete explanation of the nature of our capacity for formal understanding and for formal acts or activity. This is the best way to express our claim that neither reductive nor non-reductive accounts of the formal determinacy of understanding can constitute an intelligible explanatory framework for the phenomenon in question.

One of the distinctive features of this research project is that it promises to bring to the fore a novel case against naturalism, not by means of 1 or 2 as it is often done (for another useful collection of essays that manifests this trend, see O'Connor et al., 2003; Velmans and Schneider,

2007), but by means of recasting the insuperable difficulties that naturalism faces in the face of the formal aspects of mathematical and logical understanding. Unlike some of the most known arguments against computational theories of 3 (Lucas, 1961, 1984, 1984, 1996; for an early but influential objection to Lucas's argument, see Lewis, 1969, 1979; for a famous variant of the Gödelian argument, see Penrose, 1996a, 1999, 2011a; for a critique of Penrose, see Chalmers, 1995b; for a reply to Chalmers and others, see Penrose, 1996b), KRA is an argument against computationalism-cum-naturalism. Furthermore, our research project helps ground the claim that even if naturalizing the mind with respect to 1 and 2 is possible, it is not possible with respect to 3. Naturalizing formal understanding is in principle untenable (contra e.g., Armstrong, 1993; Dennett, 1991, 1987, 2006; Papineau, 1993, 2016; Kim, 1998, 2008; Chalmers, 2011). If so, then we need an alternative metaphysical framework to naturalism. But if naturalism faces insurmountable problems in explaining 1 and 2 then our research project will help ground the stronger claim that naturalizing the mind faces problems on *all* fronts, that is, as a plausible explanatory framework for making intelligible the phenomena picked out by 1, 2 *and* 3. Either way, KRA promises to offer an original case against naturalism *via* Kripke's argument against functionalism and computationalism and their naturalistic ontological commitments.

Our project focuses exclusively on the problem of rationality, primarily on those features of it that are properly called formal (e.g. validity). This makes it rather unconventional yet no less relevant for that matter. Research in the philosophy of mind has developed in such a way that the formal aspects of rationality have ended up taking the backseat to qualia and intentionality, perhaps due to the commonplace idea that digital computers already realize those features (see e.g., Preston and Bishop, 2002, pp. 36–39 for the claim that Turing solved the formal but not the semantic aspects of rationality). Furthermore, at a time when naturalizing the mind is in full force (see e.g. Milkowski and Talmont-Kaminski, 2010; Talmont-Kaminski and Miłkowski, 2013) the importance of bringing to the fore arguments like KRA that shed light on some overlooked mental features cannot be overstated. The present work offers a compelling case for thinking that our formally determinate cogitative abilities and acts (e.g., from reasoning in a deductively valid form to proving a mathematical conjecture) may not be a (wholly) physically realizable functional or computational process (contrary to the received view). In addition, it also promises to give us insight into what formal understanding and formal activity are. If our argument is right, then formal understanding is *sui generis*. The objective is to determine whether KRA offers the sort of argumentative and insightful support in favor of these conclusions. We will argue that it does.

Functionalism and/or computationalism can be challenged in several ways. [2] One way to challenge it is to argue that functionalism does not capture *all* mental phenomena. Most attacks on functionalism are of this sort and have to do with worries about functionalism's inability to accommodate the qualitative or phenomenal, subjective, and causal aspects of consciousness (Block, 1978, 1980a, cf. 1980b; Chalmers, 1995a, 1997, 2010; Kirk, 2015; see Braddon-Mitchell and Jackson, 2007, pp. 107–123). Another approach within this first form of challenge argues that functional organization is not sufficient for meaning; thus, functionalism cannot account for the semantic aspect of rationality (Searle 1980, 1990, 1992; cf. Preston and Bishop, 2002). The Gödelian argument against computationalism is also a species of this first way. A second way is to claim that computationalism and functionalism ends up attributing rationality or mentality to things that do not have it. Triviality arguments (Putnam, 1988; Searle, 1990) as well as liberality arguments (Block, 1978) are two different species of this second way of challenging computational theories of the mind. Yet all of these challenges are consistent with the tenability of naturalism under some other

[2] There is no conflation here. In chapter 2, it will become clear why I'm using 'functionalism' and 'computationalism' interchangeably.

(that is, non-functionalist or non-computationalist) conceptual framework. For example, one might like Chalmers (see e.g. Chalmers, 1995a, 1997, 2010) and Penrose reject computationalism about consciousness (though on entirely different grounds) and yet provide an alternative naturalistic approach to consciousness (unlike Penrose, Chalmers believes computationalism *does* at least succeed in the explanation of cognitive functions and capacities like our capacity for computing mathematical and logical functions).

A third but hitherto unexplored way of arguing against functionalism and computationalism about our cogitative powers[3] is to point out that the fundamental presupposition that a physical structure can determinately realize a definite abstract function or description is false. Saul Kripke formulated this attack during the early 1980's in a series of anti-functionalist lectures (Kripke, 1984) that have never been published (cf. Buechner, 2011). Yet his argument promises to constitute an original and serious attack not only on functionalism and computationalism but on any materialist explanation, whether reductive or non-reductive, of the nature of formal understanding. Hence, what makes Kripke's argument so important is that it seems to have far greater consequences than just the refutation of functionalist and computationalist theories of (the capacity for) formal understanding. Kripke's lectures provide the most important premise for an argument against the possibility of naturalizing formal understanding, that is to say, against the possibility of explaining in purely materialist terms the constitutive and genetic (see chapter 1 for the technical use of these terms) problem of the nature of (the capacity for) formal understanding. This is what I have called 'The Kripke-Ross Argument Against Naturalizing Formal Understanding' (KRA, for short).

Kripke's unpublished argument delivers the first premise of our argument: that no physical process or mechanism can, on account of its physical properties alone, *intrinsically, uniquely,* and *determinately* realize definite abstract functions, even such elementary mathematical operations as addition. Hence, the claim that the physical is, intrinsically and objectively, formally indeterminate. The second premise comes from a relatively unknown paper published by American philosopher James F. Ross (Ross, 1992, cf. 2008 ch. 6) wherein it is argued that some understanding and thinking intrinsically, uniquely, and determinately realize abstract functions or formal operations and therefore count, intrinsically and objectively, as real bearers of formal features. In other words, human understanding *is* formally determinate. KRA is the result of having brought together Kripke's and Ross's arguments into a single, full-blown attack on naturalism. But given the fact that Kripke's lectures remain unpublished and Ross's paper largely ignored, KRA itself remains an unexplored challenge to the prospect of naturalizing, particularly *via* computational frameworks, formal understanding and activity.

Putting together these independently developed arguments into a systematic and formidable argument has proven to be a difficult and taxing task, though immensely rewarding. One reason for this is that Kripke's anti-functionalist argument as well as Ross's argument remain relatively unknown. Therefore, there is almost no literature on either argument. Aside from the articles by Edward Stabler (Stabler, 1987)and Jeff Buechner (Buechner, 2011), and the unpublished article by Kevin J. Lande (Lande, 2011), there is no literature on Kripke's anti-computationalist argument. I suppose that this is due, in large part, to the fact that Kripke's series of lectures on functionalism and computationalism remain unpublished. Likewise, aside from some articles by Edward Feser (Feser, 2013, 2017, cf. 2009, 2006, 2014), Peter Dillard (Dillard, 2014, 2011), and Victor Reppert (Reppert, 2009), there has been no wider discussion of and published work on Ross's argument. In the case of Ross's argument, this has had the unfortunate consequence of making Feser's (to my estimation)

[3] I will use 'cogitative powers' or 'the cogitative' as interchangeable with our capacity for formal understanding, and/or formal understanding as such, and/or formal activity. It might be useful to consider 'the cogitative' as an umbrella term encompassing the capacity and the ability for formal activity, and therefore the activity as well.

incorrect interpretation and appropriation of Ross's argument somewhat authoritative, with the likes of Dillard and Reppert following Feser in their interpretation of the argument (though Dillard was the first to publish an article on Ross's argument). We shall point out the flaws in Feser's interpretation of Ross in due course (chapter 4). Many of the objections I have raised in this monograph to KRA are mostly of my own making together with those that have come about either from conversations with colleagues and supervisors or from extrapolating objections raised in entirely different contexts against entirely different arguments to our argument, KRA. Secondly, because a proper treatment of the nature of formal understanding and activity (a species of rule-governed understanding and activity) must take into account and examine the quite difficult and vast debates over the nature and varieties of computation in cognitive science, the nature and the variety of rules (formal rules in particular), the nature of formal practices like mathematics, logic, and formal linguistics, the role the concept of understanding plays in the rather perplexing sea of formal practices, the nature of the variable component of a formal structure (see chapter 1), and so on. I have not been able to do it all, and surely many mistakes and misunderstandings will be found throughout. But a dissertation must be submitted, and one cannot keep writing without end about things one could effectively keep writing about without end. I have, unlike Kripke, a penchant tendency for repetitiveness and, like Kripke, a penchant tendency for excessively long footnotes. I have done my best to rid this work of both of these tendencies, but I'm afraid I have not entirely succeeded. In a process of elimination, much has been deleted. Yet some rather long footnotes remain which would have made the chapters even longer if placed as part of the main text, creating also unnecessary disruption in the flow of the main text, and would've deprived the reader of illuminating explanations had these been deleted. As with anything, choices had to be made. Keeping some of these long footnotes might make for a rather uncomfortable reading, but I have opted for depth and detail (perhaps excessive detail) over readability. Hopefully, this choice has not come at the reader's peril.

On to the structure of the dissertation. The first chapter (The Conceptual Framework) lays out the conceptual framework of the whole work. The concepts of 'formal understanding' and 'pure formal structures' are thoroughly explained, together with the complex web of concepts within which formal understanding can only make sense. So, we analyze the notions of 'structure,' 'pure,' 'rule,' 'norm,' 'form/formal,' 'capacity,' 'ability,' 'activity,' 'practice,', and more. We mention and sketch the ways in which the capacity for language and linguistic understanding could be interconnected with formal practices and formal understanding. Kripke's argument against computationalism and for the formal indeterminacy of the physical is introduced in chapter 2 (Kripke and the Formal Indeterminacy of the Physical). We begin with (a rather free interpretation of) the skeptical paradox and work our way down to the anti-computationalist argument. Here we introduce Stabler's and Lande's own formulations of the argument and use them to strengthen our case. The third chapter (The Kripkean Challenge: Objections and Replies) is parasitic on the second chapter since it deals with possible objections to Kripke's argument and with our replies to these objections. We also outline, though only in sketch, the ways in which the physical can intelligibly be said to compute mathematical and logical functions. We attempt to show that the formal indeterminacy of the physical does not make physical computation of mathematical and logical functions impossible so long as it is properly understood in which way the physical can be said to carry out formal (mathematical and logical) operations. An asymmetry will emerge between the intelligibility of physical computation and the intelligibility of 'human computation' (formal activity), which will be the subject of chapter 4.

Having established the formal indeterminacy of the physical, chapter 4 (Ross and The Formal Determinacy of Understanding) introduces, develops, and defends Ross's argument for the formal determinacy of human understanding. We also point out some notable misunderstandings of

Ross's argument, most noticeably at the hands of Feser and Dillard. We emphasize Ross's focus on the formal aspects as opposed to the semantic aspects (content) of the understanding, point out how Feser *et al.* have conflated those aspects in their representation of Ross's argument, and briefly sketch a parallel argument as to content that is closer to Ross's original formulation of the argument. The following chapter (chapter 5: Objections and Replies) is, like chapter 3, parasitic on the previous chapter as it deals with possible objections to Ross's argument in particular, and to KRA more generally, and with our replies to these objections. The last and final chapter (chapter 6: KRA and Linguistic Understanding) extends the line of reasoning developed in the first five chapters to the problem of explaining (the capacity for) linguistic understanding and the nature of the grammar of natural languages. We focus exclusively on Chomsky's computational approach to our language capacity. This chapter reads more like an appendix, as it is neither exhaustive nor definitive. We address what I call 'Chomsky's conjunction' and go on to argue that KRA poses a dilemma for Chomsky: Chomsky's conjunction cannot be coherently sustained, so either both conjuncts are false or one of them is. Either way, the philosophical framework that undergirds Chomsky's explanatory project for the explanation of our capacity for language faces a deep problem. The chapter is not an exercise in linguistics and should therefore not be confused for a refutation of Chomskyan linguistics. The chapter is, rather, a minute exercise in the philosophical foundations of cognitive science and linguistics and does no more than pose a challenge or a set of difficulties to Chomsky's philosophical approach to the nature of our capacity for language and linguistic understanding.

As has already been stressed, however, it has not been possible to do it all. Some notable omissions deserve mention. We have not considered all or even most of the varieties (of models) of computation and explanatory frameworks within the cognitive sciences. The reader will notice, for example, that we have not addressed connectionist models at all (Aydede, 1997; Cummins and Schwarz, 1991; for an early critique of connectionism, see Fodor and Pylyshyn, 1988; for a general introduction to connectionism, see Garson, 2018). There are two main reasons for this. First, because incursion into the connectionist debate and developing Kripke's argument with respect to connectionist models would have resulted in a much longer exposition of Kripke's argument than was necessary for present purposes. Indeed, as the reader will appreciate, we formulated the argument only with respect to finite state automata (FSA), leaving it at an intuitive level its application to Chalmers's 'combinatorial-state automata' (CSA) and other relevant models of concrete computation. And even then, our discussion of FSA's is surely non-technical and quite rudimentary. Secondly, because it seemed to me that the considerations given in chapters 2 and 3 suffice to make it intuitively clear how our argument can be extended *mutatis mutandis* to any connectionist model that may attempt to *explain* formal understanding and activity in terms of the physical realization of abstract neural networks that (can) compute mathematical and logical functions. Formal indeterminacy will abound regardless if the determination of the formal properties of a concrete computation is taken to be wholly material. Thirdly, because of considerations of simplicity. We have tried to stay as close as possible to Kripke in the *type* of examples used to make his argument. Indeed, I'm afraid our use of FSA's was not necessary at all, but we've followed Kripke in utilizing FSAs to bring home the point of the argument. The argument takes place at such level of generality, at such level of fundamentality, that it affects any explanatory framework that presupposes that the gap between abstract (mathematical and logical) structures and functions, on the one hand, and physical structures and functions, on the other, can be objectively and determinately bridged by the physical. KRA shows this to be impossible. Admittedly, we have had to gloss over a host of issues that would've made for an even longer dissertation had we delved into a detailed analysis of them (e.g., an in-depth analysis of the notion of computable functions, of formal understanding *vis-à-vis* the non-computable, of issues concerning the nature of logical consequence and logical form, a more detailed analysis of program implementation and the generation of

programming languages, the relevance or irrelevance of non-classical logics for KRA, information theory, etc.). We hope, however, to have done enough to show how KRA might fare with respect to such issues and topics.

Perhaps it worth noting that the scope and field of vision of the present monograph when originally conceived was much narrower than what it is. A project that was originally conceived to consist in a rigorous and in-depth examination of a hitherto relatively unknown argument soon ballooned into an argument whose considerations inevitably overlapped with a host of problems that lie at the intersection or the union of different intellectual disciplines like the philosophy of mathematics, philosophy of science, philosophical explorations into computational implementation and computer science, the nature of artificial intelligence, linguistics and philosophy of linguistics, philosophy of psychology, and so on (viz., the philosophical and conceptual foundations of the cognitive sciences). Pragmatic and simplicity considerations have prevailed in the choice and manner of exposition of the concepts and problems. I believe that we have succeeded nevertheless in doing justice both to KRA and to the host of problems with which KRA is interconnected, despite the limited and 'in-sketch' manner of the discussion of many of these problems in the present work. We do not claim to have exhausted the field of inquiry; on the contrary, we claim that the exact opposite is the case. But KRA is a good place to start from, and we hope this dissertation encourages further research and philosophical discussion. Lastly, let me note that I propose KRA in much the same way as Davidson (Davidson, 1970) proposed his now (in)famous 'anomalous monism'. I have given a rigorous defense of KRA and have argued in detail for its soundness. Whether KRA is in fact a sound argument remains to be seen and much work needs to be done before we can assess holistically its overall cogency and philosophical significance. Only time will tell. Still, it is worthy of serious consideration and I hope that in the ensuing dissertation I have, if nothing else, managed to convince you that it does indeed merit *your* serious philosophical scrutiny and attention.

1
The Conceptual Framework: Formal Structures and Formal Understanding

1. THREE MAIN CLAIMS

The present investigation consists, broadly speaking, in the examination and defense of the following claims:

1. that Kripke's unpublished argument against functionalism (Kripke 1984; Buechner 2011) establishes (or *can* be used to establish) the *formal* indeterminacy of the physical—i.e., that no physical system, state, or process can *in itself* determinately and intrinsically realize a definite *pure function or formal operation* (whether this be a mathematical function like addition or a pattern of inference like modus ponens)—and, therefore, the computational and mathematical indeterminacy of the physical;
2. that Kripke's argument is *sound*; and
3. that the argument that results from conjoining Kripke's argument with an argument by James Ross (Ross, 1992; Ross, 2008) for the formal determinacy of human understanding is likewise *sound*.

The resultant argument I have called "The Kripke-Ross Argument Against Naturalizing Formal Understanding" (KRA, for short). The thesis to be defended is that formal operations (e.g., following or applying mathematical and logical rules), which are constitutive of human formal understanding and activity, cannot be determinately realized by a physical property, process, or mechanism; but they are determinately realized by human formal understanding. Therefore, human formal understanding cannot be wholly physical.[4] This settles part of the answer to the constitutive problem of rationality, by showing what it at least *cannot* be. But what is the constitutive problem of rationality?

2. THE CONSTITUTIVE AND THE GENETIC PROBLEM

Any account of the nature of our rational or cogitative capacity (otherwise called 'the intellect')––in particular, the capacity for formal understanding and the sort of activity that issues forth from it (e.g., doing mathematics)—must at least make clear the criteria of intelligibility that any purported *constitutive* and *genetic* explanations of the capacity in question must adhere to if they are to make sense. As the American philosopher Thomas Nagel (Nagel, 2012, p. 54) tells us, any account that

[4] There is an important distinction between 'a capacity', 'an ability', and 'an activity' at play here. Roughly, 'understanding' is constituted by ability, not capacity (except as a necessary condition). Capacities are things we are endowed with on account of the beings we are, not something which is acquired. Abilities, though grounded and made possible by capacities, are acquired. We all have in principle the capacity to swim, for example, but not all have acquired the ability to swim, and if you have not the ability, you cannot swim (i.e. you cannot engage in the activity, which is the ability exercised). The ability has to be acquired. Something analogous is the case with formal understanding. Of course, if something is the sort of being that is not capable of acquiring a set of abilities, then it cannot be the sort of thing that engages in the activity that consists in the exercise of those abilities. Likewise, if something on account of what it is cannot (come to) possess a set of abilities, then it is the sort of thing that cannot possess the necessary capacity for (acquisition and possession of) that sort of abilities. This will be, more or less, the nature of our argument against naturalizing the capacity and ability (or abilities) for formal activity.

aims to explain life, consciousness, and/or rationality as an integral part of our cosmos must satisfy two elements: first, an "ahistorical constitutive account of how certain complex physical systems are also mental", and second, "a historical account of how such systems arose in the universe from its beginnings". Let us call the first one the constitutive problem; the second the genetic problem. And let us rephrase them in the following way,

(1) **The Constitutive Problem**. The problem of explaining the nature of (the capacity for) human formal understanding. It corresponds to the question "What is *it*?"
(2) **The Genetic Problem**. The problem of explaining how the *capacity* for such understanding and how the understanding itself came about and could come about in the universe. It corresponds to the question, "How is *it* possible?" or "How could *it* have come about?"

Though the proper purview of KRA is (1), KRA does not provide a complete answer to (1) but only a partial one. A complete (or, should I say, a *more* complete) answer to (1) goes way beyond the scope of the present monograph.[5] Yet KRA does provide an answer to (1), however incomplete, that is sufficient for the purposes of a metaphysical account of our rational capacities. Secondly, insofar as an answer to (1) affects the delineation of possible solutions to (2), our argument can be said to illuminate (2). But we do not address the genetic problem in any direct way. Of course, if our argument is correct then certain philosophical accounts will be ruled out as plausible solutions to the genetic problem (e.g., materialist/physicalist accounts). As Nagel observes, an answer to the genetic problem will depend, in part, on the answer given to the constitutive problem.

Constitutive questions are complicated, and we do not hope to answer them completely. We do, however, think to have found an argument that effectively refutes and excludes a type of solution to the constitutive problem of formal understanding, namely the physicalist solution—be it reductive or non-reductive. A physicalist response to the constitutive question "what is it to compute a mathematical function," or "what is it to reason in accord with logical rules", or "what is it to have, and what is the nature of, the capacity to do mathematics and logic", for example, does not meet the criteria of intelligibility for a constitutive (and genetic) explanation of mathematical (and logical) activity. We show that formal understanding and formal *activity* (understood as the exercise of formal understanding) cannot be wholly a matter of matter—say, a matter of the physical/material realization or implementation of formal rules and structures.

3. THE UPSHOT OF THE ARGUMENT

[5] For example, a more complete account of the *capacity* must consider human nature (that is, species-specific traits) and the role human biology plays in the having and fulfillment of that capacity. Our argument rules out biolog*ism* (the materialist view that all human capacities are wholly reducible to, or no more than a consequence of, the specific biology or neurobiology of human beings) as an intelligible solution to (1), but it does *not* rule out biology from figuring as part of a complete solution to (1). A complete solution to (1) will be interdisciplinary. On the other hand, the fulfillment of that capacity (say, in the acquisition of formal understanding, of a set of abilities with respect to formal concepts and operations) and the coming about of the sort of activity which is the exercise of formal understanding (e.g., doing mathematics) requires an examination of the role normative *practices* like mathematics, which are shared and 'intersubjective', play in the development and possibility of that sort of understanding. So a philosophical investigation into the nature of normative practices and of practice-bound understanding, which such practices make feasible and relying on (e.g., insights) for further development, together with the sociology and anthropology of such practices, must figure in a complete explanation of (1) insofar as we aim to explain constitutively not merely the *capacity* but also the *ability* (the understanding, which is acquired) and the *activity*. This is all beyond the scope of this monograph, but I hope to pursue these considerations in future research.

If Kripke's argument is indeed sound, it will force us to re-examine our accounts of the nature of human intellectual activity if it is the case that *that* sort of activity genuinely realizes formal operations. Ross's argument for the formal determinacy of thought shows, in conjunction with Kripke's argument, that human intellectual activity does indeed realize formal operations and therefore cannot be (wholly) material. It is in this sense that we say that formal understanding cannot be naturalized. (Something is 'naturalizable' if its genetic and constitutive explanations lie wholly in physical/material processes and events.)[6] My research draws out some of the most important (and bound to be controversial) philosophical implications of KRA. In particular

(1) that there are no pure formal (e.g. mathematical and logical) operations or functions in matter;
(2) that human intellectual activity if it genuinely realizes formal operations cannot be (wholly) material; or contrapositively, if our intellectual activity is wholly material, then we never genuinely realize any logical or mathematical operation, and
(3) that, by extension, if linguistic abilities are partly a species of, or constitutive of, formal understanding (i.e. that the ability to follow pure formal rules is grounded and made possible by the ability to follow grammatical rules) then the pure formal aspects of language and linguistic understanding, i.e. of linguistic abilities and linguistic activity, cannot be (wholly) material either.[7]

Before laying out the propaedeutic to our argument in favor of the impossibility of naturalizing formal understanding, it is important to note that this is not an exegetical work. Of course, any work dealing with the arguments of another philosopher must manifest some level of exegetical sensitivity, on pain of attributing to that philosopher views and claims he never held. I believe this had been achieved in the present investigation. Although my examination of Kripke's argument goes in many respects well-beyond things Kripke explicitly said in his anti-functionalist lectures, the

---

[6] Let me note that in a different sense of the term I am a full-blown naturalist, if by this we mean that the human beings belong essentially to the natural order and therefore have their proper place and scientific explanation in nature. Human beings are animals, and as such a full explanation of their nature cannot do without natural and empirical (i.e. chemical, genetic, biological, psychological, anthropological, social and cultural, etc.) explanations. There is no escaping this. Yet, due in part to the nature of formal understanding and activity, one has to acknowledge that there is a capacity in these sorts of animals that transcends the capabilities of matter at any level. Yet I do not believe that the human person can be identified with this capacity. The cogitative (rationality, the intellect) is a constitutive capacity of the human person, to be sure; but it is *not* the human person. The human person is the *whole* animal, the whole *rational animal*, and cannot be identified with any of its parts or capacities on pain of a mereological fallacy. Since, however, the rational capacity (the intellect) is constitutive of what it is to be human, if KRA is on the right, there is something about the nature of the human person that cannot, constitutively and genetically, be accounted for materially.

[7] This ought not to be confused with the claim that linguistic abilities and linguistic activity have *no* neurobiological (and social) necessary grounding. I'm talking about the *formal* aspects of language which have to be part of the linguistic abilities of any competent language speaker and which cannot be explained nor reduced *in toto* to a physical or material ground. But this is compatible with the claim that there cannot such a thing as human language (at least as we know it) if not for the particular anatomical and neurobiological constitution of the *homo sapiens* and its essential and unique status as a social, culturally embedded, animal. On a separate note, let me clarify my use of the terms "physical" and "material". I use "material" and "physical" synonymously throughout. Thus, by "physical" I do not mean merely "matter at the level at which fundamental Physics studies it" but matter at *any* level, e.g. at the level of chemistry, biology, neurology, and so forth. If one takes there to be "a most generic species of material substances" (cf. Ellis 2001, 2014), namely fundamental physical particles—in the sense of being that which nothing composed of matter can fail to possess if it is to be material—then our argument will entail that nothing entirely composed of such particles can be a suitable subject of formal understanding, which rules out any purely material thing. Of course, there could be beings that are essentially material but that nevertheless possess formal understanding (viz., humans); but then they would not possess such an ability (and capacity) on account of their materiality alone. See chapters 4 and 5.

claims I'll be defending can be readily shown to follow from, and to be consistent with, Kripke's anti-functionalist argument. My purpose, at any rate, has been ultimately to go *beyond* Kripke and Ross, cashing out some of the implications of their arguments that they may have failed to properly recognize. To paraphrase Kripke, the present work should be thought of as expounding Kripke's (and Ross's) argument as it struck Antonio, as it presented a problem for him.[8]

4. SETTING THE STAGE: THE CONCEPTUAL FRAMEWORK

Let us turn now to the conceptual landscape of the present work. It is of paramount importance to be clear on how the terms (e.g. 'form', 'structure', 'pure', 'understanding', etc.) in the investigation to follow are used. The terms that I will be employing in developing KRA are rather unorthodox, and are not entirely Kripke's or Ross's, but I believe that this divergence causes no harm. One fascinating feature of Kripke's argument is that it can be fully elaborated with respect to what I call *pure formal structures* (what Ross called *pure functions* or, more generally, *pure forms*). I will explain in turn each aspect of that expression.

4.1 STRUCTURE

First and foremost is the concept of 'structure'. Now, 'structure' is spoken of in many ways. We adopt Koslicki's (Koslicki, 2008) characterization of *structure* for the purposes of our research.[9] The Koslicki account takes its cue from different sources. One way in which to begin to understand what is meant by "structure" is to consider a structure as something ("a sort of entity") that makes "slots" available to be filled by some other objects, as long as those objects "satisfy certain type restrictions"—presumably imposed by the *sort* of structure in question. As a result of filling in and occupying those slots, the objects will exhibit a certain configuration or arrangement. Now, the structure determines which *type* of objects can occupy the available slots. So long as the objects are of the right type, a number of suitable transformations (perhaps a potential infinity of them, depending on the *kind* of structure) with regards to *which* objects of the right type occupy the slots can take place without any change in the configuration or arrangement. As Koslicki (Koslicki, 2008, pp. 235–236) herself tells us,

> …structures are precisely the sorts of entities which make available positions or places for other objects to occupy, provided that these occupants satisfy the type restrictions imposed by the structure on the positions in question; as a result of occupying these positions, the objects in question will exhibit a particular configuration or arrangement imposed on them by the structure. […] What lies at the heart of the notion of structure is ultimately the distinction […] between what is taken as *variable* in a given domain and what is taken as *invariable,* relative to a set of admissible transformations.

Koslicki is sensitive to the fact that in some context structures can be spoken of *entitatively* in the sense of being objects for which proper names stand for, while in other contexts they can be spoken of *positionally and relationally* in the sense of being relations that determine the position of

---

[8] "I suspect - for reasons that will become clearer later - that to attempt to present Wittgenstein's argument precisely is to some extent to falsify it. Probably many of my formulations and recastings of the argument are done in a way Wittgenstein would not himself approve. So the present paper should be thought of as expounding neither Wittgenstein's argument nor Kripke's: rather Wittgenstein's argument as it struck Kripke, as it presented a problem for him."(Kripke, 1982, p. 5)

[9] Koslicki's account of 'structure' builds on (Harte, 2002; Nicholas Rescher and Paul Oppenheim, 1955; Nozick, 2001)

certain objects (Koslicki, 2008, pp. 252-254). Be this as it may, the notion of position at play here need not be understood as 'physical position'. For position can be quite abstract, as is the position of prime numbers in the list of all natural numbers or the position of '2' and '3' in the ordered pairs ⟨2, 3⟩ and ⟨3, 2⟩. In one sense, structures signify the presence of ordering principles, of complex patterns into which suitable phenomena can be placed so as to be ordered, or patterned, or formed, in determinate ways, giving rise (in some cases) to something entirely new, such as a configured or structured whole with features or powers that cannot be found in the component parts apart from being thus organized or arranged.[10]

What is of relevance to us is Koslicki's general characterization of 'structure' as *that which remains or is held invariable through admissible permutations or transformations of its variable component.*[11] As Koslicki points out, particular domains like logic, mathematics, linguistics, and chemistry, for example, in which the notion of structure figures prominently, "tend to aim at formulating laws whose purpose is to characterize the behavior specifically of those features of the domain which are held invariant under structure-preserving transformations."[12] Generally speaking, the structure of something is its invariable component, and particular domains cash out that invariable component and structure-preserving transformation differently, in accord with the nature of their subject matter. David Oderberg's (Oderberg, 2014, pp. 166–167) summary of Koslicki's account deserves to be quoted in full:

> The structure of something […] is not identical to but is paradigmatically expressed or defined by a set of rules (axioms, principles, further definitions) that specify the type and configuration of the elements that make up the content that bears the structure. This content is variable; as long as the elements of the content are of the right type and in the right places (both specified by the structural rules), a structured whole of a certain kind will always be present. […] Syntactic structures are given by a set of rules specifying which linguistic elements occupy which places within a linguistic sequence such that a well-formed, that is, grammatical, sequence is the result. The simple noun-verb-noun structure contains three slots that can be occupied (as the description makes plain) only by certain kinds of content element if

---

[10] In some structures the variable component itself, i.e. the organized phenomena, when organized in some ways exhibits properties or powers that it could not otherwise have. See (Koslicki, 2008, pp. 236-239) for a helpful overview of the different but interconnected grammatical roles the word 'structure' plays in language.

[11] "To illustrate, relative to the structure imposed on the dinner guests by the "alternate-by-gender" seating arrangement [male-female-male-female], for example, the particular men and women occupying these slots are taken as variable, i.e., their numerical identity is irrelevant to the question of whether the seating arrangement in question has been successfully implemented, as long as the distribution of their gender with respect to one another remains fixed. Thus, all "transformations", which involve only women switching places with other women and men switching places with other men, are admissible in the context at hand, in the sense that they leave the seating arrangement in question intact: such "transformations" result in scenarios which, from the point of view of the structure, look indistinguishable." (Koslicki, 2008, p. 236)

[12] "The numerical identity of those elements, on the other hand, which are taken as variable within a given domain, tends to be irrelevant to the laws governing these domains; the only features of these variable elements which are relevant to the laws in question are their type as well as their distribution relative to the structural elements that are taken as invariant in a given context." As we said at the beginning, "structure" is spoken of in many ways. Koslicki takes the notion of "form" to be the closest cognate to "structure" and goes on to point out some of the many ways in which "structure" is spoken of. For example, we've already seen the notion of structure as that which makes slots, or places, or positions, or nodes available to be occupied a certain type of objects under a set of admissible transformations. Likewise, structure as "configuration", "arrangement", or "pattern", has also made it into our portrayal. For still further cognates of structure in different domains and contexts, see (Koslicki, 2008).

> a well-formed sequence is to be produced. "Dogs like books", "Books like dogs", "Jim loves Jane" are all well-formed precisely because the slots in the structure are filled in the right way. A sequence, of course, can be well formed syntactically without expressing something semantically proper. […] We can see from these sorts of examples the way in which Koslicki understands the structure/content distinction. Indeed, her explanation is plausible as a general account of structure and derives its inspiration from the work of Rescher and Oppenheim, who identify the three elements of (i) parts, (ii) domain of positions, and (iii) assignment of parts to positions as the constituents of a structured whole. On this analysis, we should understand the parts as the content and identify the structure itself with the domain and assignment, the domain to be thought of as the slots or places for the parts, and the assignment as representable by a set of rules, principles, definitions, and so on.

We adopt Koslikci's account of structure but employ it in a slightly different way. We are interested in the notion of structure only insofar as it applies to and can be predicated of *formal operations,* i.e. the application and computation of formal rules and functions. More precisely, we limit ourselves to the structural principles of *formal (e.g. logical and mathematical) understanding and activity*, that is those ways in which one's thinking must be formed or configured for one to carry out a definite formal operation of the relevant sort. Our study is not a study into the metaphysics of structures, but into the metaphysics of those formal structures that constitute formal activity whether expressed in terms of the application or following of formal rules, the computation of pure functions, or the carrying out of formal operations. These are things we (can) *do* and *comprehend.*[13]

4.2 FORMAL

The notion that is more liable to cause confusion is, I think, the notion of *form*. Let us say that "form" signifies the kind of *invariable* component (the so-called "skeleton" or "abstract pattern"[14]) found in inferences, arguments, calculations, grammatically well-formed expressions, and

---

[13] Though I use 'functions', 'rules', and 'operations' in the case of formal structures interchangeably, I recognize that this can be misleading. There are computable and non-computable functions, and thus a distinction between those functions for which there can be an effective, mechanical procedure (an algorithm) for computing their values and those for which there cannot be one. I am mostly concerned with computable functions, but what I say can be readily extended *mutatis mutandis* to non-computable functions. For it is a feature of any formal structure, be it the function itself or an algorithm if the function is computable, that it cannot be had determinately, whether by supervenience or by realization, by anything physical on account of physical properties alone (this is what KRA purports to prove). Similarly, though by 'formal operations' I mean the carrying out of the computation or the application of a pure function or rule, we can also grasp non-computable formal operations (e.g., the ω-rule of induction). So KRA applies to these as well. In this sense, KRA is much broader than and differs considerably from Gödelian arguments (Lucas, Penrose), for it applies to *any* pure formal (e.g., mathematical) rule or function whatsoever. KRA is, in a way, much more radical in its conclusion.

[14] By 'abstract' I do not wish to commit to any ontology of so-called 'abstract objects'. Though I do not hold to an entitative view of anything abstract and am convinced that abstract structures are the product of human intellectual activity, the argument that I will develop need not be committed to that and doesn't necessarily conflict with an entitative view of abstract phenomena, an *ante rem* or *ante rationem* ontological view of abstract objects. So allow me this minimalist characterization of 'abstract structure': a structure is 'abstract' if the invariable component is *fully* constituted, and thus fully definable, independent of any *specific* (though not any *type* of) variable component and of any *particular* materialization. (This is more or less what is meant by 'pure'.) Two simple examples of an abstract structure (in our sense) would be the addition function and the rules for quantifiers in second-order logic. Of course, some people consider the Equator an abstract object, one which cannot be *identified* with any specific physical or material property (it's variable component?) but which cannot be fully defined and specified without some reference point in matter (e.g. the

so forth. It is that which remains invariant through suitable transformations of the *variable component*, which in this case we could simply label 'the content' of the invariable component (e.g., what a logical argument is *about,* the numbers a numerical function can take as its arguments, the grammatical units that can constitute a noun phrase in a given sentence, etc.). The invariable component in turn accounts for, and grounds, the *kind* of formal structure it is (e.g., the *kind* of inference), and therefore grounds

(a) its formal (i.e. logical, mathematical, or grammatical) taxonomy,
(b) its formal properties (e.g. truth-preserving, bijective, etc.), and
(c) its identity, completeness, and existence independent of any *particular variable* component.

One can ask what constitutes the invariable component. We could add the following condition:

(d) the invariant component is constituted by a set of axioms, principles, and/or definitions relative to a set of admissible transformations in a given domain and practice.

Is (d) true of *any* formal structure? If what (d) says is that any formal structure is constituted by axiomatization, then this is just false (e.g., consider Aristotelian term logic or non-axiomatized elementary arithmetic). Some formal structures can be axiomatized, and some will be constituted by axiomatization, but not all. If, however, (d) merely expresses that such structures are constituted by virtue of a set of stated and/or agreed upon specifications—be it in terms of axioms, principles, definitions, intuitions, or what have you—of what counts as *v* and *V-ing* (where '*v*' is a given formal rule and '*V-ing*' the application of the formal rule, the formal operation) within a given practice or domain, then (d) seems much more plausible when interpreted thus. Fortunately, we do not need to settle this here. Our argument will work equally well regardless of the opinion one holds on such issues.

Characterizing the notion of 'form' and its cognates in this way might make talk about *formal structures* seem redundant, since we've likewise characterized structures as the invariable component or principle of something (see above). But the redundancy is only apparent. By adding the adjective *formal* we wish to make clear that we are talking about a specific kind of invariable component, namely that of *structure-preserving understanding*. More precisely: *truth- or correctness-preserving thinking in virtue of form* (broadly understood). It is the ability to think *in ways that remain invariable through suitable permutations or transformations of their (perhaps potentially infinitely many) variable components (i.e. contents), that are pure and abstract, and that are dispositive of (i.e. logically and normatively encompass) any possible transformation of the variable component*. Let me illustrate. Take the traditional syllogistic form AAA-1 Barbara.

All M is P
All S is M

North and South Pole of the Earth). I am not talking about these abstract phenomena, which I characterize (see main text below) as *mixed formal structures*—a rather confusing expression at best. (Ross refers to them differently, dividing them into 'obtuse abstractions/idealizations' and 'abstract particulars'.) I am interested in the ones I call 'pure' (again, see main text below). But as will become clear, both in chapter 4 and chapter 5, since even such activities as grasping, creating, discovering, defining, and specifying mixed formal structures require that we think in a way that is formally determinate, the physical will always manifest a level of formal indeterminacy even with respect to mixed formal structures like musical scores. But this is to be expected if mixed formal structures are partly the result of conventions requiring interpretative practices and standards which supply a substantial part of the relevant determinacy. For an overview of the 'abstract vs. concrete' debate, see, e.g., (Rosen 2017, Craig 2016, 2017 (forthcoming); cf. Cowling 2017.)

All S is P.

To reason in that syllogistic form is to reason in a way that is (1) determinate in structure or form, (2) truth-preserving *in virtue of* its form (not the content, i.e. the variable component), and thus (3) determinate with respect to the formal validity of any possible (actual, potential, and counterfactual) application or instance of it so long as the transformations or substitutions of the variable component remain of the right sort (e.g., universal affirmative categorical propositions).[15] Consider these two examples:

All human beings are mortal.
All Puerto Ricans are human beings.

All Puerto Ricans are mortal.

and,

All artefacts are man-made.
All guns are artifacts.

All guns are man-made.

Both are cases of AAA-1 Barbara, even though the content (i.e. what the argument is *about*) is *not* the same. The *content* of that argument is what we call its "variable component", i.e. that which can change without a change in the structure or form of the argument. But what makes them be cases of the same type of argument is their *logical form* and this is what we say remains invariable through suitable transformations of the variable component. It is the form, i.e. the invariable component, that which makes each one be the *kind* of inference it is and, thus, the determinant of their logical validity. The same can be said of disjunctive syllogism

Either α or β
Not α

β

[15] Of course, if one espouses the Boolean interpretation of categorical term logic, one would *not* grant that *any* kind of universal affirmative categorical proposition is of the appropriate type for that syllogistic form, but only *contingent* universal affirmative categorical propositions. Similarly, what is a perfectly valid form of argument in Aristotelian term logic,

All M is P
All S is M

Some S is P.

comes out as invalid in Boolean term logic (i.e. existential fallacy). But our argument is not committed to either interpretation of term logic, nor to a particular view of the nature of logic expect as something we do. Logical rules constitute ways of reasoning. But which interpretation best captures the truth or nature (if any) of logical rules is immaterial to our argument. Perhaps the Aristotelian view of term logic is false or inadequate; but we can surely reason in accord with it. This is what I am interested in. This is not an investigation in the philosophy of logic, but rather an investigation into the capacity for logical reasoning and the ability to reason logically.

and, in fact, of any other logical form. Furthermore, this is true not only of logical forms but of also of mathematical and grammatical forms. A rather straight forward example is the formula for calculating the slope of a straight line:

$$\frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

When one understands what this formula expresses, one is able to calculate the slope of a straight line (that is, the rate at which a change in the value of x produces a change in the value of y) in terms of this equation. For example, if we are given that $y = f(x) = x^3 - 2$ and that two points on its graph have the following x-coordinates: $x_1 = 4$, $x_2 = 6$, then we can determine the values of y_1 and y_2 on the basis of the function $f(x)$—after which, given the formula above, we can proceed to calculate the slope of the straight line under consideration. Here, as with logical forms, we have both an invariable (i.e. the form or structure of the operation(s))[16] and a variable component (i.e. the values of the variables). The variable component (that is, the values of the x- and y-coordinates) can be any real number, and any suitable transformation of one real number for another will not constitute a change in what is being done but only a change on what it is being done *to*. What we see above is a symbolic schema or representation of a form of calculation, i.e. a form of understanding.

In linguistics we find something analogous.[17] For the sake of simplicity, consider the famous **rewrite rules** in early generative grammars (Chomsky, 1957) which are instructions for generating the structure of grammatical phrases or sentences. According to this early paradigm, sentences

---

[16] To be sure, the formula for calculating the slope is, like many other mathematical formulae, constituted by simpler mathematical operations like addition (or subtraction) and division (or multiplication). Our investigation will reveal that unless the form of the operations is determinate *in the doing*, in the carrying out of the operations (e.g. in *adding* x- or y-coordinates), there cannot be such a thing as *calculating* the slope of a straight line, since "calculating such-and-such" is something we do, an activity, and its specification will depend both on which formal operations are carried out (the form of the doing) and over what (the content of the doing) relative to a given domain and practice. One of the virtues of our overall argument is that it is formulated using only the operations of elementary arithmetic, e.g. the addition and the identity functions. It should take almost no intellectual effort on the part of the reader to see how the same argument can be developed using other mathematical operations and functions as well as logical and grammatical ones. In fact, what we have provided thus far are symbolic-syntactic representations of forms of understanding. A calculation or an inference cannot be identified with a symbolic notation and a set of syntactic rules developed for representing and capturing it. For there are multiple ways of symbolically and syntactically capturing and representing the very same formal structure or operation. This is, I believe, readily demonstrable in the case of mathematics (e.g., unary notation as opposed to binary or 10-base notation, or squaring using Arabic numerals as opposed to Roman numerals) and logic (e.g., *modus ponens* in Polish first-order logic notation). Yet syntax is paradigmatically a formal structure (at least the syntax of artificial formal languages) so our argument will apply as well to its alleged material realizability. A separate but interconnected question is whether our argument can be extended to the grammatical structures of *natural* languages and therefore to linguistic understanding as such. This will be more controversial, but if one approaches the study of natural languages with the idea that natural languages are akin to formal *calculi*, then our argument will apply to linguistic understanding as such. Plus, if one accepts the view of natural language as a formal calculus, one will have to commit oneself to the claim that at least some grammatical forms cannot be identified with the particular symbolic-syntactic structure of a concrete natural language since multiple and diverse natural languages can contain, capture, represent, and express the very same grammatical forms (say, the same principles), albeit differently (say, a different alphabet and morpho-syntactic rules concerning order and construction of discernible grammatical units or principles) somewhat in the way *modus ponens* can be expressed and captured in different formal languages. Note that though such 'higher forms' may not be identifiable with any particular syntactic/representational system (a particular symbolic-syntactic system) they cannot exist but in a given syntactic/representational system. But a Platonist may beg to differ. Settling such disputes is not the subject of this monograph.

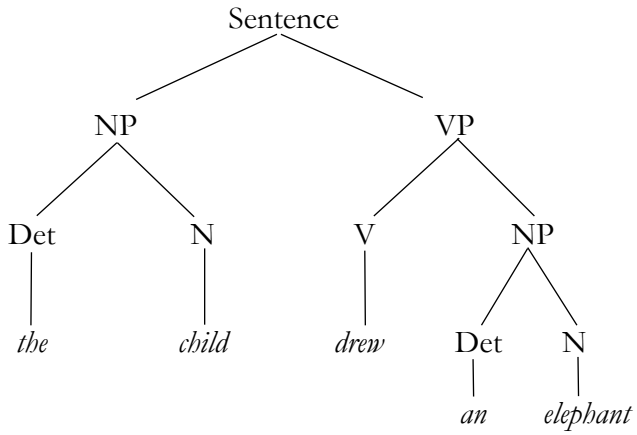
[17] In what follows, I rely heavily on (Cook and Newson, 2007).

consist of structural groupings of words (i.e. phrases) and are said to have phrase structure. Rewrite rules are taken to be instructions for generating *grammatical* phrase structures. These are quite abstract and are represented symbolically in various ways. One way is by tree diagrams—similar in appearance to truth trees in propositional and predicate logic. Take the following sentence:

(1) The child drew an elephant.

There are two phrases in this sentence: the noun phrase (NP) *The child* and the verb phrase (VP) *drew an elephant*. Like truth trees, the tree diagram of a sentence breaks its phrase structure down to its lowest, atomic level. In the case of (1), for example, the VP breaks down into a verb (V) *drew* and an NP *an elephant*, whereas the NPs in turn break down into a determiner (D or Det), *the* and *an*, and a noun (N), *child* and *elephant*, respectively. The tree diagram of (1) then looks like this:

(S)



The actual words at the bottom can be substituted for indefinitely many other words to form a new sentence with the same structure, i.e. phrase structure, so long as the word-substitutions are of the right type. The phrase structure of (1), call it (S), is repeatable and instantiable by infinitely many possible sentences. This formal structure is what the person with linguistic understanding of English knows, being that on account of which she is able to generate ever new sentences that are in accord with a grammatical structure. There are also sentences that do not possess this structure, of course, but the point of illustrating this is to show that phrase structures are the abstract invariable component of innumerable actual and potential sentences.[18] What Chomsky did was formulate a simpler and more powerful form of representation of phrase structures (that is, rewrite rules) such as:

(S) $S \rightarrow VP\ NP$

where the symbol '$\rightarrow$' stands for 'consists of', 'S' for 'sentence', and so forth. One then goes on to provide the same rewrite rule to all constituent expressions (e.g., VP and NP). Thus,

(Sa) $VP \rightarrow V\ NP$

[18] To be sure, the notion of phrase structure in linguistics goes back to the 1930s and precedes the advent of Chomsky's transformational generative grammars. I am here interpreting them in light of the 'cognitivist turn' in the late 50's due to (Chomsky, 1957, 1959).

(Sb) NP → Det N
Etc.

This is a crass simplification. The theory is much more complex and technical than this, and it is admittedly in many respects a thing of the past (transformational generative grammars have long since abandoned Chomsky's 1957 rewrite rules; in part due to Chomsky himself having replaced them for more technical and 'powerful' ones). But the general principle remains operative: that there are formal structures in language whose invariable component are the subject-matter of formal approaches to languages and whose actualization in particular sentences lies in the application of highly abstract, formal (i.e. phrase structure and transformational) rules. If this is so, then our argument will have to apply to linguistic understanding and linguistic activity (we will go briefly into Chomsky's project for linguistics in chapter 6).

The point of all of this is to illustrate that our claim is a generalizable one: a pure form or a pure formal structure can be logical, mathematical, or grammatical. The expression 'formal structures' will therefore refer, in our technical sense, to what Ross called 'pure forms' or 'pure functions', namely formal rules, functions, and operations.[19] So we are interested in the forms that constitute the structural principles of higher human intellectual activity as well as the capacity, and consequent acquired ability (or set of abilities), for that kind of activity. This includes language, as formal approaches to linguistics make manifest.[20] At any rate, we can surely say, albeit rather

---

[19] Again, whether they are *just* forms of human understanding is something I do not, and cannot, purport to argue here. The issue is beyond the scope of this investigation. But what I do defend is the claim that, if anything, pure formal structures must be at least (capable of being) forms of human understanding. Otherwise we could not do what we do and what we know we do when reasoning validly (or invalidly) or computing a mathematical function. On a slightly different note, perhaps it would be better, for our purposes, to speak of "correctness-preserving activity" or "form-preserving thinking" rather than "truth-preserving so-and-so", which might be associated exclusively with truth-functional logic. It would be odd, for example, to suggest that linguistic understanding and grammatically well-formed linguistic activity is a matter of reasoning, akin to way we reason in truth-functional logic. I am of the opinion that linguistic understanding is a condition for the possibility of any other kind of formal understanding. But this is, ultimately, irrelevant to the overall argument to be considered. The point is that there is such a thing as linguistic understanding, which consists in or begets the ability to follow certain formal rules (e.g., the morpho- and logico-syntactical combinatorial possibilities of languages).

[20] Even if you object to the very plausibility of formal linguistics (say, suppose you are staunch radical defender of sociolinguistics), you still cannot reasonably deny that we can *do* formal linguistics, that we can create, define, and formulate invariant linguistic laws or structures, follow them or apply them, and so forth. Whether there are actually any structures in natural languages of the sort formal linguistics claims to uncover or discover is of no relevance here. They may or may not exist in natural languages, but they do exist in formal languages and in formalizations of natural languages. So even if they are mere artifacts of human intellectual activity, they are still *present* whether as mere creations or not in that sort of activity; it remains that we *can* (i.e. have the capacity and acquired ability to) create, formulate, and think in accord with them (think of the invention of context-free grammars, for example). Issues about whether natural languages contain a "deep structure" that is context-free and/or computational do not need to be settled here. Our argument goes through regardless of the position one takes on such issues. For example, some opine that natural languages are not computational systems and that linguistic understanding therefore cannot be a matter of computational processes and of operating within a computational system. But this is immaterial to the issue at hand. Suppose KRA is sound. Then, if natural languages are computational systems, it will follow that the capacity for such systems cannot be purely physical (nor purely physically determined). If, however, natural languages are not computational systems, then it may turn out to be the case that the capacity for natural languages is a purely physical capacity—so long as natural languages are in no way (constituted by some other species of) pure formal structures. Suppose the latter alternative is intelligible (I have my doubts). Then all that would follow from this is that the capacity for language is not the capacity for formal understanding, that is, for following and applying pure formal rules and carrying out pure formal operations. Yet it would remain that any capacity for formal understanding could not be a purely physical capacity, and since we do have the capacity for formal understanding (e.g., for computing, for inventing and operating in accord with computational systems), it follows that at least some capacity of ours cannot be purely physical. To be sure, I do concede the premise that natural languages are constituted by pure formal structures,

imprecisely, that formal understanding consists minimally in the ability to engage in an activity that is constituted by

(i) an invariable component, or pattern, that is
(ii) pure (abstract) and
(iii) normative

Logical, mathematical, and grammatical forms all satisfy these conditions. By normative we mean nothing more than that a pure function (e.g., a formal rule) constitutes a *norm*, the standard of correctness for its application. Now, how exactly different formal structures are characterized will depend on the nature of the activity and practice that is so *formed* (inferring deductively, measuring angles, counting discreet units, arranging members in a group, parsing, etc.). For example, in standard first-order propositional logic, the logical forms are taken to be truth-functional and are taken to represent deductively valid forms of inference.[21] But this is not so for grammatical forms (e.g. phrase structure) or trigonometric forms, which are neither truth-functional nor symbolic representations of deductively valid arguments. It is not even true of all logical structures (e.g., term logic and intensional logics), a subset for which truth-functionality either doesn't hold at all or is insufficient. Our characterization of formal structures is therefore at the upmost general level, and as such does not gainsay that there are different *species* of pure forms or that some pure forms may be more fundamental than others. Our point is merely that the activity that is the exercise of formal understanding is constituted by (i), (ii), and (ii). Logical, mathematical, and grammatical forms are all species of such kind of invariable components. Chapters 2 and 3 will focus, primarily, on the *physical* realization or implementation of logical and mathematical structures. We give our argument for the computational and, more generally, mathematical indeterminacy of the physical there. Chapter 6 will tentatively extend the argument developed in the first two chapters to Chomsky's theoretical or cognitivist approach to grammatical structures and linguistic competence.

It should be clear that by 'structure-preserving understanding' I mean the sort of understanding which is acquired and manifested in the activities proper to definite formal practices and disciplines. Take high-school level trigonometry. When a pupil understands and knows what it is to solve a right triangle[22], she can proceed by thinking in terms of any one of the trigonometric

---

specifically by generative computational mechanisms *à la* Chomsky. But this last one is a dialectical and 'for the sake of argument' concession. For, in chapter 6, I will show that if this view of language is correct, then Chomsky's neurobiologism about the computational deep structure of natural languages cannot be correct. This must not be mistaken for agreement. I am no computationalist, and therefore do not agree with Chomsky's conception of a *natural* language as constituted by something akin to a formal calculus or mechanism whose formal rules are fully specifiable independently of pragmatics and (to a more limited extent) semantics, and whose rules budget normatively for all possible or conceivable applications. But this is neither here nor there. For a formidable position along these (anti-Chomskyan) lines, see (Braine, 2014).

[21] To be sure, there are disagreements about how to understand the notion of logical consequence. The two main accounts of logical consequence (though these do not exhaust the available options) are the substitutional account and the model-theoretic account (see e.g. Gamut 1991a & 1991b; Sider 2010). Again, it does not matter for our argument. We are asking a different question. Let us put it this way: if you are the sort of being that can understand first-order propositional logic and the different approaches to logical consequence, then you are the sort of being that can acquire formal understanding and, thus, the sort of being that can carry out formal operations. No being whose capacities are, constitutively and genetically, purely physical can acquire formal understanding and carry out formal operations. Hence, no purely physical being can be the sort of being that can understand first-order propositional logic and the different approaches to logical consequence. Since we do engage in such activity (see chapter 3), our formal activity and our capacity for such an activity cannot be purely physical.

[22] When a pupil comes to know, by calculating correctly, the values of all the angles and all the sides of a triangle. We call this "solving a triangle".

functions, which are forms (otherwise called 'rules') for calculating the angles and the ratio of the sides of the right triangle. Whenever she calculates the sine of an acute angle θ of a right triangle, for example, she is thinking in the form 'sin $\theta = y/r$' (where y is the side opposite θ, r the hypotenuse). The specific notation is the symbolic representation of the invariable component that structures, i.e. *informs*, her calculation and whose admissible variable component[23] can change in value without a change in what is being done. This is why we can *repeat* the same operation over ever different values. Lastly, the sine function is dispositive of (the value of) any possible argument. Whenever I competently proceed according to the form 'sin $\theta = y/r$', I think in a way that is dispositive of, and determinate with respect to, any possible case of that form. But this does not mean that in order to understand the sine function I must grasp *every possible*[24] case of that function (an impossibility, since

---

[23] I am considering the sine function as a function from $\mathbb{R}$ to $[-1, 1]$. We can of course consider it differently, say, by restricting the codomain or letting "sin" be a function from $\mathbb{R}$ to $\mathbb{R}$, in which case "sine" will represent a different function. So depending on the *choice* of the domain and codomain different functions come about. But nothing I've said undermines or is undermined by this.

[24] Though the *capacity* for formal understanding is the *capacity* (in the sense of being *capable*, as opposed to *able*) for understanding any formal rule and operation whatsoever and therefore for applying and carrying out *any* possible formal rule and operation, this does not translate into the *ability* to do these things. *In principle*, if one is the sort of being that can acquire formal understanding, one is the sort of being that can acquire complete mastery of any possible formal rule and operation. But this does not mean that therefore one will in fact acquire complete mastery and grasp of such rules, functions, and operations. What is true of capacities does not necessarily translate into abilities because the latter are acquired in practice and training, and like any ability their acquisition depends on and is affected by an array of accidental, contingent, circumstantial, and limiting factors constitutive of our being the kind of animals we are. Our higher (innate) capacities, however, do not seem to be thus constrained. Consider linguistic understanding. Our capacity for language or linguistic understanding is the capacity for any (possible) natural language (and, I opine, for any possible artificial, formal language), but none of us have the ability to speak and follow the rules of every and any possible natural language. The latter is by its very nature something acquired and learned. (Though, we say, once one has acquired mastery of a given language, e.g. one's first language, then the ability to do one thing in that language is the ability to do them all, though in actuality one only does finitely many.) There is, however, an analogous issue concerning the abilities for formal activity: whether the ability to apply a given formal rule in a given case *is* the ability to apply it in any (possible) case, or whether one has to acquire a new ability for every application of the rule. The exercise of that ability will, of course, be limited by innumerable accidental factors. But the question is whether the ability to carry out e.g. the addition function correctly for finitely many pair of numbers *is* the ability to carry out that function for *any possible*, and thus for infinitely many, pair of numbers in the domain of the function (ignoring for now accidental and limiting factors on the exercise of that ability). It seems that the answer is yes. Otherwise, one would have to acquire a new ability every time one is to apply the addition function to new or never before added numbers, and the ability to add at each stage would be indeterminate with respect to addition since it could be made out to accord and conflict with incompossible counterfactual cases. But it also seems that the answer is no. For then, two individuals who cannot add the same pair of numbers (suppose one of them possessed a prodigious memory and is able to compute greater numbers mentally than the other) would be said to have exactly the same ability, but only one of them can exercise that ability to a greater extent. But what exactly is the problem with this last statement? Couldn't we just say that the difference does not lie in the ability to add *per se* but in the ability to add *quickly* and *a lot*, that is, in greater speed and memory? That is, it seems the difference would lie not in their mastery of the function (the invariable component) but in their different abilities to process functors (the variable component) of different degrees of complexity. So the limitation concerns the "processing" of the variable component, not the mastery of the invariable component. Whatever the correct answer to these questions might be, it remains that due to features such as finite time and length, finite operative capabilities, differences in cultures and practices, in neurological makeup, etc., we are not *able* in practice to grasp and compute *any and every possible* case of a pure function (say, numbers so great that the computation would exceed the lifespan of the universe and the finite memory capabilities). But these limiting features, though essential to our nature as *animals*, seem accidental to our nature as *rational agents*, that is, accidental to the formal operations and activity themselves. Take a person with a cognitive impairment who can nevertheless do mathematics, albeit with quite considerable limitations. She learns to add and actual does add, but her ability and exercise of that ability is constrained by accidental features to that ability and activity, namely neurological limitations. Yet she really adds, and were we to remove the neurological impairments, there would be no cognitive-neurological barrier for her to reach full, standard competence. Analogously, we could say that even cognitively healthy human beings *qua* animals are nevertheless in a condition of existence that

the domain and range of that function are infinite), just as in order to understand a rule one does not need to grasp every possible application of that rule. Otherwise it will be impossible to ever understand a pure function or formal rule. [25] But it has to be the case (with qualifications) that the ability to follow and apply a given *formal* rule or to compute a *pure* function *is* the ability to follow and compute a rule or a function that is determinate with respect to any possible case of it and therefore that the ability to compute that function makes the computation of any possible argument of that function potential, though not actual.[26] The point is that to *do* a certain kind of formal activity, like calculating the sine of a right triangle or deducing one of the disjuncts of a disjunction from the negation of one of its disjuncts, is to engage in an activity that is formally structured and

---

impedes the unlimited exercise of their formal abilities which seem to be potentially infinite or limitless. This is a point similar to that made by the likes of Fodor and Chomsky to the effect that our linguistic and formal activity is potentially infinite, though actually finite. It is also similar to Aristotle's point that the human mind is potentially anything thinkable, or rather, that anything that has being is potentially intelligible to human beings though in actuality only a handful, finite sort of intelligible beings will ever be grasped by us.

[25] Note that it makes no sense to say of the sine function (or any other trigonometric function) that it *ought to* or *should* answer or produce such-and-such a value to such-and-such an argument. That's like insisting that the standard of correctness be correct (a conceptual confusion, at best). Here's a preliminary way to flesh out the notion of reasoning in a way that is determinate in structure or form. It cannot be the case that the sine of a 30° angle, where the 30° angle in question is one of the acute angles of a Euclidean right triangle, would give you or would have given you .534 if carried out Wednesdays, π if applied after 2098, or 0 if it is or had been 'carried out' by a computer program for the 58th time. For then it would not be the sine-function that is being applied. It would not be truth-preserving in virtue of the form of sine because it would not be sine-preserving, and thus would hardly count as a genuine case of the sine function (more on Chapter 2). An interesting case in trigonometry is what is called the 'ambiguous case', but even there the possible values that the sine function can take are determinable.

[26] No one can compute all the possible values of a pure function, especially if it's infinite. But in our case it is not necessary to do so. The whole function (that is, the invariable component) is contained in a single act of understanding and of genuine computation (see chapters 4 and 5). The limitations here seem to concern the variable component, not the invariable component. The fact that I cannot generate a sentence, recursively, that is made up of 2 billion sub-clauses does not mean that I cannot generate a sentence, recursively, that is made up of sub-clauses or that I could not, in principle, generate a sentence that long if things like fatigue, memory loss, lack of patience, were removed. It seems like I do have, in principle, the ability to do *that* insofar as I have mastery of the formal rule that generates such types of sentences. But for that I need not be able to bring about all the possible cases that the application of that rule could generate. It's like trying to do arithmetic with roman numerals or reason syllogistically in Chinese. It's a matter of the functors, not of the function. Problems of logical omniscience do not follow, for one can be said to know something without knowing it completely or exhaustively and one can be said to have mastery and expert knowledge of something without thereby knowing everything that could be logically constructed as following from it. It just isn't necessary. Otherwise, any expert in any field would be omniscient. Now, by saying that the ability to compute certain values of a pure function makes the computation of any possible argument of that function potential just means that once the ability is acquired there is, *in principle*, no limit to the extent to which that ability can be exercised or further developed to compute ever new values of the sort of function in question. There *are* limits to what we can actually do, of course, such as limited life-span, a limited amount of energy and concentration, finite memory, cognitive limitations, etc. But these are *accidental* to these sorts of operations and constitute practical and extrinsic rather than constitutive constraints on such abilities and operations. Again, this is another difference between purely material (e.g., biological) capacities/abilities and formal capacities/abilities, which are embodied. (Again, this is a similar point to made by the likes of Aristotle, Chomsky, and Fodor, regarding the potential, not actual, infinity of operations that intellectual and language-possessing beings can engage in.) The qualifications are the following. Since formal understanding is *at least in part* ability-like, there are degrees of understanding and therefore there is such a thing as partial as opposed to complete mastery of a relevant formal technique(s) (cf. Hacker 2005, p. 18). This means that if one's understanding of a given formal rule or operation is partial (to be sure, partial understanding itself constitutes a spectrum), then one is not able to grasp any possible application of the rule or operation. But one is *in principle* able to grasp any possible application of that which is understood. So if one understands the addition function partially, say because one does not know that there's addition over negative integers, the person will (ignoring accidental factors like time, length, tiredness, cognitive impediments, et.) be in a position to grasp and apply any possible application of the addition function over positive integers (assuming that one's understanding of that function is limited to its application over positive integers). See footnote 24.

thus determinate with respect to form—a form which allows for the decidability and correctness of any counterfactual claims about that formal activity. The same goes for many other elementary mathematical and logical functions (chapters 2, 3, and 4) and, perhaps, for grammatical functions (chapter 6).

4.3 A CONCEPTUAL DISTINCTON: FORMAL AND SEMANTIC (INTENTIONAL) ASPECTS

An important distinction, fundamental to the argument developed in this monograph, is that between the *formal* (i.e. structural) and the *semantic* (i.e. intentional) aspects of understanding. Though I do not hold that it is possible to have one without the other *in re*, as they are two specifications of the very same capacity (like two sides of the same coin, so to speak), their conceptual differences are not always respected, and philosophers tend more often than not to conflate them.[27] The differences can be pointed out like this. We do not merely think or wonder *about* pure functions (e.g. thinking about addition or the square root function or conjunction). We actually think *in terms of*, or *compute,* or *carry out* these functions.[28] In other words, we *add*, *square*, *conjoin*, and so forth. We *do* these things, we do not merely contemplate them. We really reason validly (and invalidly), we do not merely think about valid or invalid ways of reasoning. And so on. When we (competently) think *about* the logical form of 'conjunction', *that* logical form is the intentional aspect (i.e. the content) of our thought or judgement (the form being the structured way in which it is being thought about conjunction, say "If a conjunction is true, then its conjuncts are true. This conjunction is true. Therefore, its conjuncts are true."). But whenever we *conjoin*, then the logical form of conjunction is the formal aspect (i.e. the structure) of our thinking or reasoning (the content, if any, being whatever the propositions or expressions conjoined are about). This distinction does not entail that therefore one can competently think about or understand a formal rule without being able to genuinely follow or apply it, nor that, conversely, one can follow or apply a formal rule without being able to think about it or without understanding it. I side with Peter Hacker (Hacker and Baker, 2014, 2005, 1984) in saying that the ability to do one presupposes and entails the ability to do the other (this will become clear when we come to such things as defining, explaining, formulating, proving, etc., specific formal structures; for, to do such things one must be able to think in a way that is both semantically and formally determinate). Our investigation, however, will focus primarily on just one side of the coin: the formal side. It is the ability to *follow pure formal rules and carry out formal operations* that is the focus of our investigation. We are interested primarily in the nature of the *act* and *activity*, that is, in such things adding, squaring, inferring, calculating, solving, proving, etc., and the ability (or set of abilities) whose exercise counts as doing these things.

Detour: one can think about many things one does not grasp or comprehend, just as many people can think about the Big Bang theory without any understanding of it. (A person who understands nothing of the Big Bang theory can still think about it even if by means of some sort of propositional mediation of the sort 'that theory first proposed by a Belgian priest', 'that theory which proves the beginning of the universe', 'something talked about in cosmology', etc.). Thus, one can think about a formal rule without understanding the formal rule. But this is not the sort of

---

[27] The literature on Kripke's rule-following considerations is vast and still growing, and there are too many substantial differences in the responses to Kripke's skeptical argument. But the conflation of the formal with the semantic or intentional aspects of rule following seem to be a constant throughout (see e.g. Kusch, 2006; Boghossian, 1989; cf. McDowell, 1984)

[28] We also attribute such verbs as 'performing', 'carrying out', 'executing', etc., to computers, but they apply to them only analogously and derivatively, the way a robot is said to be dancing or the way auto pilot is said to fly, and not univocally with what we do. This will be made clear in the arguments laid out in the following chapters.

'thinking about' that I have in mind. By 'thinking *about* something' I mean thinking with a grasp as to what is being thought about. This I called 'thinking competently about *x*', which is just another way of describing one's thinking about something which one understands. When you understand something, you can think competently about what is understood. Of course, your competence will be proportional to your degree of understanding, but the point is that to think competently about *x* you have to have (some) understanding of *x*. Now, if one understands some formal operation *v*, one knows what it is to carry out *v* and possesses the ability to *V* (that is, can engage in *V*-ing). This is, I take, trivial. But it seems that one cannot say the same about the ability: that in order to be able (i.e. in order to possess the ability) to carry out *v* one must understand *v* (I presume that "in order to carry out *v* one must be able to carry out *v*" is incontrovertible). One thing is to conceive of a scenario in which one carries out *v* without the ability to *V* (e.g., by accident, luck, cheating), another is to say that one can have the ability to *v* without any understanding of *v* (e.g., the way digital or analog computers compute). In regard to the former (following a rule without being able to follow the rule) I would simply say such a thing makes no sense and *cannot* count as a case of following/applying any rule at all. It is a case of mere conformity or coincidence with the application of the formal rule rather than an instance of the application of the formal rule. In contemporary jargon, it is a case of *as if* following or applying the rule rather than a case of following/applying the rule. As to the latter (being able to follow a formal rule without understanding the formal rule) it could be a formidable counter-observation if it were the case that there are some formal rules one could apply without understanding them.

Consider: "Don't people reason fallaciously all the time? Do not most people reason, unbeknownst to them, in accord with logically invalid forms of argumentation? Aren't they *able* to do this? Yet they do not understand what it is to reason invalidly or fallaciously, or at least what it is to reason in accord with logically invalid forms. People are in general not very good at reasoning and many are ignorant of this fact. Hence the ability to reason in accord with a logical form does not require or imply any understanding of that logical form." It is doubtful whether people under normal conditions do not really understand their reasoning.[29] Note that there are different kinds of fallacies, most notably *formal* fallacies and *informal* fallacies. The former arise out of the *form* of the reasoning or argument, the latter out of the *content* of the reasoning or argument in everyday discourse. Informal fallacies are, perhaps, the most common ones and are part of the subject matter of what has come to be known as 'informal logic'.[30] Yet it remains that in order to engage in the sort of activity that constitutes the subject matter of informal logic one must be linguistically competent

[29] Recall that 'understanding' is, at the very least, ability-like. Nothing excludes that people may understand their reasoning imperfectly or incompletely and that therefore they are able to reason thus-and-thus without understanding *why* they are reasoning thus-and-thus (e.g. their motives, biases, etc.) and without being able to tell that their reasoning so-and-so is not a valid form of reasoning—in the logical sense of 'valid' (e.g. truth-preserving). Understanding is gradual and completable. It comes in degrees, like abilities, and can be perfected, sharpened, or deepened.

[30] "Most fallacies, however, are not formal but *in*formal: They are patterns of mistake that are made in the everyday uses of language. Informal fallacies, which we examine very closely in this chapter, arise from confusions concerning the *content* of the language used. There is no limit to the variety of forms in which that content may appear, and thus informal fallacies are often more difficult to detect than formal ones. It is language that deceives us here; we may be tricked by inferences that seem plausible on the surface but that are in reality not warranted. Such traps, the "pitfalls" that language sets, can be avoided if the *patterns* of those mistakes are well understood. Considerable attention will be devoted to these informal fallacies—the kinds of mistakes made in everyday speaking and writing, and commonly encountered, for example, in the "letters to the editor" in daily newspapers." (Copi et al., 2014) Note the claim that natural language deceives us and is therefore unreliable for rigorous logic, all of which motivates the idea that natural language is unfit for rigorous logical thinking and analysis and therefore an artificial formal language is needed to capture the hidden or disguised logical structure of thought—one which avoids the "pitfalls" or "traps" of natural language. This is contentious, but it was and still is one of the most appealed grounds for justifying the turn from natural language to formal-mathematical languages in logic.

in a natural language and be able to engage in the sort of informal reasoning that informal logic studies and analyses. One's natural language here serves as the condition for the possibility of, and the medium-in-which for, any inference. Sure, one can commit a formal fallacy while reasoning informally (e.g., like committing the fallacy of affirming the consequent while arguing with friends about some political issue at a local bar) just as one can make a mistake at calculating the total cost of what one has picked up at a grocery store, and one can reason in a formally valid way while reasoning informally (e.g., like reasoning in accord with *modus ponens* while debating about a controversial supreme court ruling with one's siblings at a family dinner) just like one can calculate correctly the total cost of one's grocery before paying. Fallacies are *mistakes* of reasoning of an interesting sort and therefore parasitic, or unintelligible without, the standard(s) of logical correctness from which they deviate.

Now, being able to detect that one's *form* of reasoning is invalid or that one's inference on the basis of the *content* of a given set of premises or assumptions is fallacious is not a necessary condition for reasoning invalidly or fallaciously nor for *understanding* the form or content of the reasoning which is invalid or fallacious.[31] Otherwise everyone who reasons invalidly or fallaciously would do so willingly and knowingly, which is in fact not the case. But in order to follow whatever the reasoning might be and to carry it out one must be in a position to understand *it*. When people under normal conditions (e.g., no cognitive malfunctions) reason invalidly without knowing it, there *is* an ability and there *is* understanding: the ability in question is the ability to *do* something, namely, to reason in a determinate way that, unbeknownst to the reasoner, is logically defective. It does not follow from the fact that she doesn't *detect* the logical incorrectness of her thinking that she does not understand her reasoning or that she cannot reason fallaciously since she doesn't know that she is reasoning fallaciously. One thing is to know how to reason or what it is to reason in a certain way, another is to know how to reason *well* or that that way of reasoning is a *good* way of reasoning. All that follows from her ignorance is that she does not understand *all* of the aspects of the reasoning she is engaged in, say that some things as opposed to others do or do not follow from such type of reasoning. This is compatible with what we've said thus far. Otherwise it is hard to see how one can speak of an ability to *V* or as that person having the ability to *V* if there is no grasp as to what it is to *V*. There must be some aspect under which *V*-ing is understood in order for one to be able to *V* (again, I am only talking of such activities as are determined by formal concepts and rules, like in the application logical, mathematical, or grammatical rules—the grammar of natural languages creates its own problems, but it is still the case that any competent speaker of a language understands the rules or 'principles' of the language under various aspects and to different degrees). I am prone to think that natural languages, in a host of different ways and to different degrees depending on the natural language in question, already contain logical structures as constitutive elements of the language such that the rules of natural languages are logico-grammatical and therefore any competent speaker of a natural language will be able to *reason* (make inferences, say) in logically determinate ways and apply (and misapply) to whatever extent the 'informal' (but 'formal' in our sense, and 'formalizable' in the

[31] Here's an exercise for logic teachers and tutors: include, as part of the particular proof system one is going to use to introduce one's students to the propositional calculus, the rule of affirming the consequent and the rule of denying the antecedent (both of which are formal fallacies) as part of the rules of inference of the formal language of the propositional calculus. You will include them as if they were sound logical rules of the calculus, and so will not inform the students about the fact that these are in fact formally invalid ways of inference. The pupil, if a total novice, will come to understand the rules and learn what it is to apply either one. She will reason in accord with those rules, and thus will have the ability to affirm the consequent and deny the antecedent (assuming she didn't have it already). She will understand these rules. Yet she will not know that these are invalid forms of reasoning until she is told (say, by the tutor) or somehow discovers this to be so. So, again, one can understand what it is to *V* without thereby understanding or knowing all there is to *V-ing*. Yet the *form* will be determinate regardless.

mathematician's sense) logico-grammatical rules and structures that she knows as a competent language speaker.[32]

Natural languages might also contain the basis for the possibility of mathematical understanding at a higher level, like containing primitive or advanced ways (say, by the evolution of the language in a culture where modern mathematics is finely engrained) to signify quantities be them discrete, continuous, and/or to generate an indefinite number (or a potentially infinity according to Chomsky) of discreet grammatical units. At any rate, our argument need not depend on the biconditional statement 'one understands *v if and only if* one has the ability to *V*' (where '*v*' is any pure formal rule); the conditional 'if one understands *v*, then one has the ability to *V*' would suffice. For, all that KRA purports to show is that the physical does not and cannot have the ability to compute, apply, or carry out any *v* and therefore cannot be a proper subject of formal understanding. I would, however, assert the converse if weakened as follows: if one has (or can have) the ability to *V*, then one (is the sort of being that) *can* understand *v*'. Put this way the converse of the original conditional statement is no longer (as) contentious. Let me note an additional problem. Unless the activity in question is the exercise of understanding (abilities), there is no differentiation between actually carrying out *v* (that is, following and applying the rule *v*, i.e. *V*-ing) and merely conforming with *v* (i.e. behavior that is only *as if V*-ing behavior or behavior that is interpretable as and that counts in a derivative-dependent manner as *V*-ing in virtue of the prior, real human activity of *V*-ing).[33] There would be no intelligibility to the different between *V*-ing and *as if V*-ing—something demonstrably false. Phrased differently: if one knows what it is to carry out *v* and possesses the ability to apply *v*, then one understands *v*. Thus, with respect to formal operations and rules, we can state the following:

> Take any pure formal rule, call it ψ, then we can say that one understands and can engage in Ψ-ing *if and only if* one knows what it is to apply or follow ψ and possesses the ability to apply or follow ψ.

This is not, however, generalizable to every kind of *doing* or activity. There are various acts or activities (e.g. 'squatting' in weight-lighting, 'stealing a base' in baseball, 'performing a quadruple lutz' in Olympic figure skating, etc.) whose understanding does not beget the ability to do them, though it does beget some sort of ability (or abilities) with respect to them (say, explaining which acts satisfy the relevant criteria for being a case of such activities, identifying and pointing examples of them, spotting errors or mistakes in interpretation or application, etc.). Formal rules and operations are slightly different in that their understanding consists, partly, in being able to apply them and carry them out. One cannot be said to understand addition and yet not know how to add, and one cannot be said to understand the algorithm 'carry and borrow' without knowing how to 'carry and borrow' (see my footnote on Lonergan concerning two different ways of speaking of 'understanding'). Their understanding begets the ability to apply or carry them out. But this is not so with many other kinds of activities—including, of course, many formal ones (e.g., one cannot apply the ω-rule of induction because it requires an infinite number of applications) though understanding these begets certain other formal abilities (e.g. formulate useful *schemata* for the ω-rule of induction, recognition of its

---

[32] Needless to say, different languages can express the same logical rules but in grammatically different ways. Some similar is true of different formal (e.g. programming) languages.

[33] The distinction at play here is analogous to the one between derived and underived intentionality, and between underived and *as if* intentionality (see Haugeland, 1981, cf. 1997; Searle, 1980, 1992, 1983; Fodor, 1989). In general, we I talk about derived formal determination, I mean either one of the latter two: either derived formal determination or *as if* formal determination. We will come back to this in chapter 4.

inadequacy for Peano Arithmetic or any effectively axiomatized formal system, and so forth). This observation points towards the claim that there is no one-size-fits-all (i.e. univocal) analysis or definition of 'understanding', nor of the host of interrelated notions (e.g. 'ability', 'act', 'activity', 'practice', 'knowledge', and so forth). If you are a Wittgensteinian or an Aristotelian (or a weird blend of both, as I am), then this is as it should be given the nature of language, practices, forms of life, and being. If, however, someone remains unconvinced by our general observation that the *act* of *V*-ing (as opposed to *as if V*-ing) is an exercise of, and consequent upon, an understanding of *v* and that understanding *v* consists, partly, in being able to (carry out, follow, or apply) *v*, then it would be best to ignore our insistence on the concept of understanding altogether and rehash the overall argument merely in terms of *capacities*, *abilities*, and *activities (*e.g. the ability to *v* and the activity of *V*-ing).[34] For what our investigation proves is at bottom that, for any *V*-ing, if *V*-ing is a kind of formal activity (e.g. carrying out a formal operation, following and applying a formal rule, etc.) then *V*-ing cannot be either constitutively or genetically (causally) wholly physical; therefore, neither can the ability (or abilities) to carry out or apply *v* be wholly physical since *V*-ing *qua* act is the exercise of that ability.

I should perhaps point out that nothing thus far commits us to the claim that the subject matter of formal sciences like pure mathematics are mere forms of human understanding. It does not exclude it either. We here follow Bernard Lonergan (Lonergan, 1992). Paraphrasing him, we note two features of pure forms:

> In the first place, they are brought about and constituted by the mere activity of supposing, thinking, reasoning, considering, formulating, defining, and so forth. They may or may not be more than that. But if they are more, then they are not merely forms of understanding. And if they are no more than supposed or considered or thought about or followed or performed or done, still that is enough to constitute them as forms of understanding. In the second place, pure forms do not occur at random; they emerge in thinking, reasoning, supposing, considering, defining, formulating, and so forth; and that many-named activity occurs, not at random, but in conjunction with an act of insight. [35]

As the passage makes clear, our approach is minimal: whether pure forms are more than merely forms of human understanding, we neither consider nor exclude, but that they are at least forms of understanding we both hold and defend. Likewise, nothing in what we've said warrants the claim that mathematical structures are reducible to logical structures, or that all formal structures are nothing more than grammatical structures *à la* Wittgenstein. We leave the latter open. The present investigation is *not* an investigation into the philosophies of mathematics and language but into what used to be called *rational psychology*, and even here we are concerned only with formal activity and practices, variously described as 'following formal rules', 'computing pure functions', 'carrying out formal operations', 'defining and specifying formal structures', and the like. It is a strength of the

---

[34] Unless otherwise stated, '*v*' stands for an arbitrarily selected pure formal rule or function. The capitalized rendition (i.e. *V*-ing) expresses the *act* of applying, following, or carrying a formal rule (i.e. *v*). I use the letter '*v*' merely to express or represent the point that these concern things we *do* and therefore grammatically expressible in terms of *verbs*.

[35] The original quote concerns concepts, e.g. geometric concepts of a line and a point. "Note, then, two properties of concepts. In the first place, they are constituted by the mere activity of supposing, thinking, considering, formulating, defining. They may or may not be more than that. But if they are more, then they are not merely concepts. And if they are no more than supposed or considered or thought about, still that is enough to constitute them as concepts. In the second place, concepts do not occur at random; they emerge in thinking, supposing, considering, defining, formulating; and that many-named activity occurs, not at random, but in conjunctions with an act of insight." (Lonergan, 1992, pp. 32–33)

argument to be developed that it need not take sides on many such contentious, metaphysical issues. Ours is an investigation into the nature of a *capacity*, *ability*, and *activity* (understood as ability exercised): the capacity for acquiring the ability (or abilities) to think in formally determinate ways and engage in formally determinate activities and practices (like trying to prove Goldbach's conjecture). Again, consider the ability to reason deductively, that is, in ways that are truth-preserving in virtue of form for any possible case of the form, as in the case of the logical forms of standard first-order logic. Whether first-order logic concerns mind-independent or mind-dependent structures, for example, is not immediately relevant. For regardless of that metaphysical dispute it remains the case that we *do* and *can* understand the pure logical forms of first-order logic and therefore (can) follow and apply the relevant formal rules—which is why, amongst other things, we can prove its consistency and its completeness and that it will satisfy non-standard models. The same goes for the natural capacity and the acquired ability to *do* arithmetic, algebra, trigonometry, calculus, number theory, abstract algebra, graph theory, topology, and so forth. So, whatever the subject-matter of mathematics and logic is, and whatever its metaphysical status, it remains that we *do* logic and mathematics, and thus that we have the capacity and, with training and practice, the (acquired) ability to think in logically and mathematically determinate ways.[36] Something analogous is true of language and linguistic understanding

It is, however, not possible to study such an ability without paying attention to and considering its exercise: the operations. For we have no other way of studying and coming to understand the nature of an ability, and the underlying capacity for acquiring that ability, except by attending to its exercise and the practices or activities in which it is paradigmatically acquired and manifested. Thus, though the capacity to reason logically and mathematically ontologically precedes our reasoning thus (and, as such, precedes the abilities and practices in which we thus reason), it is epistemically posterior to it such that we only come to know the nature of the capacity by coming into contact with the acquisition and exercise of the ability in question.[37] In our investigation, we mostly focused on rather simple cases of following elementary formal rules and formal operations, such as adding rational numbers, squaring a positive integer or thinking in terms of the formulae for

---

[36] This should not be confused with the claim that therefore logical propositions and mathematical propositions ought to be treated univocally, as if they were the same kinds of propositions. Again, the claim being made here is only that reasoning logically and mathematically displays, depends on, and is constituted by the ability to think in formally determinate ways. Any other human activity that displays, depends on, and is constituted, either in whole or in part, by such an ability will be subject to whatever extent to the argument and conclusions of our present investigation. This should come as no surprise, since formal reasoning is in a sense rule-governed reasoning, and logical and mathematical reasoning are two special kinds of rule-governed reasoning which intersect and overlap considerably. But this does not entail that, therefore, logical propositions and mathematical propositions should be accounted for univocally and uniformly, any more than the practices of mastering *bel canto* and the tenor saxophone ought to be accounted for univocally and uniformly simply because they display, require, and are constituted in part by the ability to breath diaphragmatically. The analogy is admittedly imperfect (i.e. you can (try to) play the saxophone or sing bel canto without diaphragmatic breathing but you can't (try to) *do* logic and mathematics, that is, reason logically and mathematically, without thinking in formally determinate ways; that makes no sense. At best, it would be a case not of *doing* but of *simulating* what one does when one does mathematics or logic). But it brings home the point that the practices of logic and mathematics depend as a necessary condition (or, in Kantian terms, as "a condition for the possibility of") both on our having the capacity for formal understanding and the acquired ability to structure our intellectual activity in formally determinate ways, analogous to the way the practices of *bel canto* and saxophone playing depend as a necessary condition both on the capacity for (diaphragmatic) breathing and the acquired ability to breath diaphragmatically.

[37] Charlers Taylor (Taylor, 1964, pp. 6-8) objects (and rightly so) to the idea that purposeful or intentional behavior is behavior caused by some sort of entity (e.g. purpose or intention) antecedent and independent of the behavior which causes and explains the behavior. The entity in question is an invisible or imperceptible hypothetical entity whose only evidence is the occurrence of the behavior that its operation is supposed to cause and explain. We agree with Taylor. This is conceptually awry. It is not our position. See our remarks on behavior in the conclusion.

calculating π, but the applicability of our considerations to higher mathematical and logical functions should be pretty straightforward. It is left to the relevant experts on each field to apply our schema to the operations proper to those higher, pure formal sciences.

Two last observations seem in order. First, though ours is not an inquiry into the philosophies of mathematics, language, and logic, we do not wish to claim that our investigation will have absolutely no bearing on philosophical accounts of the ontology and epistemology of formal concepts. For one of the consequences of our argument is that pure forms *qua* rules, functions, or operations are (indeed, *have* to be) intrinsically, non-derivatively, and constitutively *present* in our reasoning whenever we genuinely reason in terms of any such logical or mathematical forms. It cannot be by extrinsic and derivative realization (that is, by instantiation or exemplification of something *else,* something extrinsic to the reasoning), for reasons to be spelled out in chapters 2 and 3. Logical forms, for example, *constitute* ways of reasoning; they are not superadded to the reasoning nor are they something extrinsically and relationally had by the reasoning itself. Thus, any philosophical account of the nature of pure forms that entails, or requires, that they be extrinsically, derivatively, non-constitutively and, therefore, relationally present in our reasoning cannot be correct. But there may be multiple philosophical accounts of the nature of logical and mathematical forms that entail no such thing. Our research offers nothing to settle the disputes between those accounts.

Secondly, consider the following passage about a sphere in Euclidean geometry which aims to illustrate the disparity between a mathematical object and a physical object:

> Consider any point in space; call it P. P is only a point, too small for us to see, or otherwise sense. Now imagine a precise fixed distance away from P, say an inch and a half. The collection of all points that are exactly an inch and a half away from P is a sphere. The points on the sphere are, like P, too small to sense. We have no sense experience of the geometric sphere. If we tried to approximate the sphere with a physical object, say by holding up a ball with a three-inch diameter, some points on the edge of the ball would be slightly further than an inch and a half away from P, and some would be slightly closer. The sphere is a mathematically precise object. The ball is rough around the edges. In order to mark the differences between ordinary objects and mathematical objects, we often call mathematical objects "abstract objects." […] When we study geometry, the theorems we prove apply directly and exactly to mathematical objects, like our sphere, and only indirectly and approximately to physical objects, like our ball.
>
> (Marcus, 2017)

That the physical can only approximate, that is, imperfectly realize *any* geometrical figure is commonplace. But that the physical can only approximate pure functions, that is, imperfectly realize and carry out formal operations (say, '$A = \pi/4\ D^2$', calculating the area of a circle, or '$\alpha \rightarrow \beta,\ \neg\beta\ \therefore\neg\alpha$', validly inferring the negation of the antecedent of a conditional, or simple addition '$x + y = z$') is not so commonly recognized. And this is our current predicament: on the one hand it is generally acknowledged (and rightly so) that formal objects like the numbers or Euclidian geometrical figures are only approximated by physical structures, but on the other hand philosophers generally proceed as if it were different with formal operations, as if formal operations could be *really* realized and carried out, as opposed to just approximated and simulated, by physical processes and mechanisms (in particular, by physical computing mechanisms). Indeed, with the advent of the computer, it is now a given that the physical realization or implementation of formal operations is, *in principle*,

unproblematic.[38] But we hold that this is simply confused and involves a category mistake between formal operations and physical 'operations' (processes). To say that something approximates *x* is to say that though it is not *really* a case of *x*, it can be taken as such by being close enough to *x*, say, due to some relation of similitude and representational reliability (e.g., introducing, replacing, and "manipulating" symbols in ways which correspond to the agreed upon conventions about symbolic transformations). We will argue that this is the proper understanding of physical, concrete computation: an approximation, emulation, and simulation. Following and applying a formal rule, grasping and carrying out a formal operation, simply cannot be a purely physical process, however complex. Chapters 2 and 3 will take pains to dispel the conceptually confused view about the potential realizes of formal operations. [39]

Before proceeding, however, I wish to make the following proviso. I recognize that the use of 'structures' to signify 'forms of understanding' can be misleading. First, because the notion of structure generally suggests something other or more than ways or patterns of thinking and reasoning. Even in mathematics, the notion of "structure" is used to mean much more than a relevant formal rule, function, or operation (e.g., a *group* in group theory). Secondly, given that 'structure' is associated in the philosophy of mathematics with the theory known as *structuralism* (Benacerraf, 1965, 1973, 1998; for a useful survey of structuralism in the philosophy of mathematics, see Horsten, 2018; Shapiro, 2018), it might make it seem that I am a structuralist or at least squeezing in some form of structuralism into the conceptual framework of the argument. Things are not so. As I say below, substitute if you wish 'structure' for 'form' or 'rule' or 'function', and the argument will work equally well. The variable component of a formal structure (where, again, by 'formal structure' we mean a Koslicki-structure that is a pure norm that structures cogitative acts or thoughts for some end(s) determined by the practice in question, e.g. 'reasoning in a way that is truth preserving', 'determining the cardinality of the set of all prime numbers', 'multiplying real numbers', etc.) structure could be an *object* of some sort. Perhaps Benacerraf's (anti-Platonist) argument against any non-structuralist, entitative view, of numbers is true.[40] Our argument is compatible with this, as we are not concerned with mathematical *objects* but rather with mathematical *operations*. Likewise, whether numbers are constituted, and can be reduced to, the abstract structural relations that are found to obtain when we specify and concoct certain mathematical operations is something about which our argument has nothing to say. 'Rules' and 'operations' of the relevant sort, that's all we are concerned here with. Thirdly, because 'structure' can in another sense of the word be taken to connote an *object* of some kind. We do not wish to commit ourselves here to such a position. In fact, it is immaterial to the argument whether one holds to an entitative, realist view or to a non-entitative, anti-realist view of pure formal structures. Our argument cuts through the divide between the view of mathematical statements as statements that pick out definite existent abstract *objects* and the view of mathematical statements as statements that do not pick out any existent abstract objects,

---

[38] *In practice* it is highly problematic, but not *in principle*. The 'hard' or 'utmost difficult problems' about human psychology lie elsewhere (see e.g. Chalmers, 1996, 1995a, 1997, 2012; Dretske, 1988, 1981; Levine, 2001; Fodor, 1975, 2008, but see 2000).

[39] We make no claim as to the metaphysical status of formal 'objects' like numbers, geometric figures, sets, groups, etc. Whether they are mere constructs or conventions, or grammatical rules (Wittgenstein) we do not consider here. Again, such issues are not only beyond the scope of this investigation, but also not to the point. Recall that we are interested in those pure formal structures that constitute the invariable components of what we *do* when we reason, deduce, calculate, measure, etc. Whether numbers are no more than abstractions from material entities (see e.g. Franklin, 2014; cf. Oderberg, 2007, p. 9), the product of creative intellectual activity, or what have you, is the proper subject matter of an altogether different kind of investigation.

[40] Our argument is compatible with structuralist (cf. Benacerraf, 1965, 1973, 1996, 1998) but also with non-structuralist approaches to numbers and set-theoretic definitions of functions. Our argument might, however, be a problem for full-blown Platonism accounts of formal (mathematical and logical) structures.

say, abstract descriptions lacking referent (see e.g. Piccinini 2015, pp. 6-8 in the context of computational objects or descriptions). The former usually goes by the name of 'realism' or 'platonism' (Balaguer, 2018, 2016, 2001, 1995; Rosen, 2018), the latter by the name of 'fictionalism' (Field, 1980, p. 189, 1998, 1989; Paseau, 2016). If you take pure formal structures to be no more than pure formal *descriptions*, which refer and describe putative abstract objects which in reality do not exist, our argument will be unaffected so long as the fictionalist does not take formal operations to be a species of such putative formal objects. For, even the fictionalist needs to explain *our* ability to do mathematics, including our ability to engage in formally determinate cogitative acts like determinately describing, abstractly, such putative but non-existent mathematical and logical entities.

Be that as it may, if KRA is sound, it would remain not only that no physical system or mechanism can uniquely and determinately satisfy a definite pure formal description, but that no physical system can on account of physical properties alone structure its "acts" in such a way as to count as the sort of thing that can provide determinate descriptions of the sort the fictionalist claims have no referent (that is, formally determinate descriptions and definitions not compossible with any other formally divergent descriptions and definitions, see chapter 2 for more). At any rate, it is hard to see how one can be a fictionalist about formal *operations*, and I am not sure anyone is. For that would seem to entail that though we describe certain activities as mathematical and logical (e.g., 'simplifying a radical,' 'solving a differential equation,' 'multiplying quaternions,' 'applying the rules for truth-trees to molecular sentences of the propositional calculus,' 'deducing by *modus tollens*,' 'deciding whether a given formula holds in a given frame,' and so on) by means of their pure invariable component these activities aren't *really* mathematical or logical since there is no such thing as mathematical and logical operations in any real sense. We do not pursue this imagined fictionalist view about formal operations any further, but we shall come to close to it when refuting the intelligibility of applying Dennett's instrumentalism about intentionality to formal understanding, viz., to the formal aspects of human computation (chapter 4).

4.4 PURE

I have already hinted at what the adjective 'pure' is supposed to connote. I do not, however, take the adjective 'pure' to be coextensive or synonymous with the adjective 'abstract', since not everything that can be said to be picked out by the latter can be said to be picked out by the former, although everything picked out by 'pure' can be said to be 'abstract'. It is, of course, not an easy task to characterize or define such an adjective, nor to draw a precise line. And any characterization I give of what counts as 'abstract' is bound to be contentious. But since this term is already in currency within mathematics (e.g., pure mathematics), we adopt and adapt it. The adjective 'pure' here is no more mysterious and spooky than when used in mathematics, or logic. It usually picks out those mathematical concepts and structures that either have no applicability to actual material reality and therefore are purely abstract (and so are their proofs, definitions, axioms, etc.) or are studied without any concern to their applicability.[41] We maintain this connotation in our use of 'pure structures', but expand it to cover *all* mathematical structures, and all formal structures in general (e.g., logical structures). Applied mathematics is in the business of finding, or creating, pure formal structures that can be used to engineer, model, or manipulate physical phenomena. An interesting question is why some pure formal structures can be applied while others cannot. We will briefly

[41] For a useful illustration on the ambiguity of the concept of mathematics, see (see Hacking, 2014, cf. 1979 for a similar treatment of logic). For a fascinating explorations on the notions of 'proof' and 'progress' in mathematics from a mathematician's perspective, see the seminal paper by (William, 1994).

touch on this in chapter 3, but only in a sketch. Our understanding of mathematics and logic as pure will become clearer when we state it explicitly (see a) below).

Objection: "one might say that since not all formal structures need be formal operations and, therefore, need not be reducible to formal operations, you have not proven that *no* pure formal structure is be physically realizable or implementable after all. Your argument if sound will only show this to be the case with respect to a minute subset of the totality of pure formal structures." Fine. I "minute" subset will suffice, for the following reasons. First, it would still remain the case that formal operations are not physically realizable in a determinate way (if KRA is sound). So, again, our argument is fundamentally unaffected. We do not need to hold that *all* formal structures are not physically realizable (though we do). Indeed, perhaps some formal operations are not even humanly realizable. This is all well and good. Our argument concerns those that are and shows that no physical thing can in virtue of its physical properties alone be said to realize them in the same sense in which we 'realize' them in formal understanding and activity. But secondly, suppose it turns out that pure formal structures (e.g., fields, groups, finite-state automata, etc.) have formal operations constitutively such that the relevant formal operations are essential to the structures, then such structure could not be the structures they are if not for the formal operations that, even if partly, constitute them. Then this lends considerable support to the stronger claim that pure formal structures in general (not merely formal rules and operations) cannot be physically realized and implemented in any intrinsic and mind-independent way. I think this is indeed the case but KRA leaves this open. It is important to be clear on what this implies and on what it does not. It does not imply that there is no such thing as concrete physical computation nor, more generally, that there is no such thing as physical satisfaction, exemplification, or realization of pure formal structures. Concrete physical computation of a pure function *is* possible, just as physical satisfaction or realization in general of pure formal structures (e.g., $E = mc^2$) *is* possible. But one must investigate in which way *that is* possible (alternatively, in which sense is it intelligible to say that something physical can realize, implement, or compute a given pure formal structure). There *is* a fact of the matter (see e.g., Piccinini, 2015, pp. 11-12) as to which computations are performed by suitable physical computing mechanisms (chapter 3, §7), but this fact of the matter is not something determined *by the physical properties and mechanism(s) of the system alone* (chapter 2). We show the way in which the physical cannot intelligibly be said to realize pure functions, namely in an intrinsic, determinate, and non-relational (or non-relative) way. Incidentally, it turns out that this is not true of human formal understanding, whose exercise constitutes formal activity—an activity that can and does realize pure functions in an intrinsic, determinate, and non-relational way—, and therefore an asymmetry between the physical and the understanding emerges. This, again, is the topic of chapters 4 and 5.

Recall our caveat about using 'structures' to denote formal operations, rules and functions, their applications and computations. Throughout the present monograph, I have kept my talk of 'formal structures' to a minimum in order to avoid some confusion, preferring instead such expressions as 'pure forms', 'pure functions', 'pure formal rules', and 'pure formal operations', all of which are more readily seen to be about things we can do (e.g., apply a formal rule, compute a function, etc.). I have not, however, rid myself altogether of the expression 'pure formal structure', and this is because I take formal understanding to be in part the ability (or set of abilities) to structure one's activity in formally determinate ways.

Now, what is a pure function? Consider the following characterization:

> "A pure function is one satisfied by an infinite number of functors, and determinate for each n-tuple, like "+", and complete in each and every instance (that is, it does not consist in a relationship among its cases, but explains them)." (Ross, 2008, p. 197 n.9)

Admittedly, this is not a strict definition but a *metaphysical* or *ontological* characterization, yet one that is satisfied in general by infinitary mathematical functions. With a tweak here and there, one can show that Ross's characterization extends to finitary functions. But what exactly do we mean by 'function' here? For our purposes, we can consider mathematical and logical functions as *rules* that determine operations (e.g., mapping) over certain determinate sort of objects (e.g. natural numbers, sets, topological groups, etc.). If this is of no satisfaction to someone who wishes to keep a strict separation between functions and rules, or functions and operations, then we can consider pure functions as those pure, abstract structures (in the sense of Koslicki-structure) whose concrete instantiation consists in the application of a set of formal rules (e.g., effective procedures, algorithms) to admissible variable components (e.g., numbers, propositions) that results in an action that is, structurally, formally determinate (e.g., a formally determinate effective procedure for some computable function).[42] This seems to tilt us towards issues of computation and effective procedures. As any mathematician would tell you, computability theory is but a branch of mathematics. A very important branch, of course, but a branch nonetheless. Computation does not exhaust the actual and possible realm of formal (mathematical) understanding. But a feature of computational theories of the mind is that the idea that *physical* computation can capture and provide a framework in which to potentially explain *our* ability to compute many mathematical and logical functions in a naturalistic way. Kripke's own argument against computationalism as well as Ross's argument for the formal determinacy of thought is formulated using elementary arithmetical functions like the addition and the identity functions, which are computable. So we will, as was previously suggested, follow suit and formulate KRA in terms of computable functions and our capacity for acquiring the ability to apply the formally effective procedures that reliably lead to the solution of a mathematical or logical problem. So long as KRA is not taken to be restricted to issues surrounding computability and effective procedures but rather applicable to mathematical and logical activity as such, such a methodological restriction should not be confused for a metaphysical-ontological restriction.

Note that effective procedures, in the sense relevant for computing mathematical and logical functions, are formal rules that prescribe a certain course of action or activity (cogitative, primarily). The action or activity that obtains as a consequence of the application of the formal rules must itself be formally determinate, otherwise indeterminacy issues of the sort we will raise (chapters 2 and 3) will abound and it will be impossible for there to be, determinately and objectively, a case of an effective procedure of the relevant sort (for an interesting, though not entirely uncontroversial, account of effective procedures, algorithms, and computable functions see Cleland, 1995, 2001, 2002).[43] Pure formal rules can be said to fall under our characterization of pure formal structures just

---

[42] In the case where the concrete computation of mathematical functions is understood as *physical* computation, the instantiation or application of such rules will come down to some sort of causal-physical mechanisms that somehow apply or follow such formal rules (see e.g. Chalmers, 1994, 1996, 2011, 2012) . Though in general we argue against this, if taken literally, we show that there is a way to account for physical computation and that therefore there is intelligibility to the notion of the physical computing a mathematical function but it won't be univocal with human beings' do (or can do) when they engage in concrete computations and thus cannot count, strictly speaking, as doing mathematics (or logic). Unlike Gödelian arguments against computationalism, our argument applies equally to computable and non-computable functions (e.g. the ω-rule). No physical computing mechanism can on account of any inherent or intrinsic power and property alone be a case of a consistent axiomatized formal theory that is negation-incomplete, because it cannot a case (in the aforementioned sense) of a formally determinate structure at all! So our argument can be true, regardless of whether any Gödelian argument is. See chapter 2 for more on this.

[43] A fascinating and interesting debate is that between the conceptual interconnections between recipes, effective procedures, algorithms, and computable functions. In particular, the first three concepts have been given wide application outside of mathematics and for non-mathematical purposes (e.g., for a humorous example, see (Stewart, 2001) for an algorithm for computing the date of Easter and its graph which resembles a regular lattice with slight

as much as the very functions they are supposed to compute (see our classification of structures, in particular category a) below). This is why the rule-following considerations are relevant, since in order to think mathematically and logically formal rules have to be followed and applied (even if only in certain domains). At any rate, so long as Ross's metaphysical characterization of what a pure function is (or our characterization of what a pure formal structure is) equally characterize, at a metaphysical-ontological level, both logical and grammatical rules or functions, then our formal determinacy considerations will readily extend to logical and grammatical structures. With respect to modern logic, the applicability is obvious. With respect to the grammar of natural languages, things are a bit different; but again, recall that we are interested in a certain kind of ability and activity (e.g. the ability to do formal linguistics) rather than on whether a certain view of language (or logic or mathematics) is correct. Yet, if the grammar of natural language turns to be constituted in part by pure formal structures or if it is (that is, if being a linguistic animal or an animal capable of language) the foremost condition for the possibility of formal understanding then the capacity for language cannot be naturalized (assuming the argument to be defended in this monograph is sound).

Having said that, Ross's characterization of a pure function will be presupposed throughout, though it should be noted that it is not necessary for the purposes of our argument to restrict ourselves to infinitary functions and rules (i.e., functions with an infinite domain and range, and rules with an infinite number of possible applications). For, the argument works equally well with functions that are satisfied by an indefinitely large yet finite number of functors—the computation of which would nonetheless take longer than the life-expectancy of the universe. (Indeed, *any* finite function will do just as well for our argument. But there is something to be said about those functions with an infinite domain and range, and we shall stick, for the most part, with Ross's characterization. It is, as Ross himself says, easier to prove the argument using infinitary functions.) Again, recall that we are interested not in formal *objects* but in formal *operations* and therefore in the nature of our capacity and ability to apply formal rules to formal and (derivatively?) non-formal objects (e.g. as when we measure the tangent of a flag post).

Now, though this characterization would suffice for the purposes at hand, it seems appropriate to venture into a more precise, albeit brief, specification of the ontology of such pure functions, without entering into the metaphysical disputes about whether they are mind-independent or not. This will be a short exercise in what Aristotelians call 'Natural Philosophy'. The notions of 'pure form(s)', 'pure function(s)', 'pure formal operations', and 'pure formal rule(s)' will be used throughout interchangeably, and as the members of a subset of the set of pure formal structures. But not everything denoted by the term 'formal' is created equal, of course, since such terms are differently used.[44] For our purposes, we use the expression as follows:

---

variations in the plot points from the lattice intersections, like the arrangement atoms in a quasicrystal!). It might very well be that our argument extends to the ability to carry out any effective procedure, and even heuristics, so long as it is rule-governed activity; but we shall not pursue this line of inquiry. We'll simply state that if the argument can be generalized, then we will have to hold a non-realist view (Chapters 2 and 3) about the objective and determinate physical implementation effective procedures and a realist view (Chapters 4 and 5) of the objective and determinate cognitive or cogitative implementation of effective procedures by human beings. See the introduction of chapter 2 for a clearer statement of the realist vis-à-vis the non-realist view of physical computation. For classic technical introductions to computation, see (Davis, 1982; Davis et al., 1994; Minsky, 1967). For more on the debate on effective procedures, algorithms, and computation, besides the papers by Cleland already cited, see (Cleland, 2004, 1993; Davis, 1982; Horsten, 1995; Horsten and Welch, 2016; Israel, 2002; Seligman, 2002; Shapiro, 2001). For the *loci classici*, see (Gödel, 1934, 1931; Turing, 1936; Church, 1936).

[44] Two warnings. First, the notion of 'form' is being used more closely to (though not identical with) the way it is used in logic and mathematics nowadays, and not strictly in the Aristotelian metaphysical sense of the *essence* of things. Of course, our use of 'form' and 'formal' is not incompatible with the latter but would have to be taken as derivative from it. But this is an issue we need not develop here. Second, our notion of 'formal' is *ontological* rather than strictly logical

a) *pure formal structures*. Structures, like the Pythagorean theorem or the rule of inference of disjunctive syllogism or Turing machines, whose *identity*, *existence*, and *completeness* neither require nor depend upon any type/token physical feature or property, or (the possibility of) any kind of materialization.[45]

Alternatively, pure formal structures are those whose real definition does not include matter either explicitly, implicitly, or virtually.[46] This is the realm of pure mathematics and logic, and it may also be the realm of grammar if formal approaches to natural languages are correct. Thus, morpho-syntactic structures may fall (with some qualifications) under this kind of structure (we will have more to say about this in chapter 6). The extension of a) includes those functions and operations proper to, and constitutive of, elementary arithmetic (e.g., addition, multiplication), algebra (e.g., simplifying radicals), trigonometry (e.g., the trigonometric functions), calculus (e.g., derivatives), etc., as well as the more complex formal operations of abstract algebra, thee-valued logics, set theory, number theory, computability theory, topology, and so forth. Again, though pure formal objects like numbers and geometric spheres can be said to satisfy a), we will be concerned only with formal operations, rules, and functions. At least two additional kinds of structures can be specified,

b) *mixed structures*. Structures, like Prokofiev's *Dance of the Knights* or the Equator or a financial institution (say, an investment bank), whose identity, existence, and completeness are not dependent on any *specific* kind of material but are nonetheless dependent on, and require, some materialization.[47] Abstract particulars are a case of

---

and mathematical as in 'axiomatized formal theory'. The addition function for example would not in the logician's sense of 'formal' be considered a formal structure, but rather an arithmetical operation that can be formally captured or expressed (i.e. represented in a syntactically interpreted artificial formal language). The logician's sense of formal falls under our ontological category of 'pure formal structure', but the latter is broader including *all* operations (and structures generally) whose invariable component is pure or abstract and normative. We can call it a Koslicki-formal structure.

[45] This is not to say that the world we live in played or plays no role in the kinds of formal structures that have come to constitute particular formal sciences (e.g., geometry). However pure formal structures might have come about, the end result is structures that belong to an altogether different ontological category than that of material structures to the effect that what can be intelligibly said of the latter cannot intelligibly be said of the former. There are no fundamental particles in, or a subatomic level to, a formal system (say, Robinson's Q arithmetic). There is no such thing as the atomic or quantum level of algebra. The formal is on another ontological plane or category. Note that to say that there are structures whose identity, existence, and completeness does not depend on materialization is not to say that those structures are not physically or materially realizable. Philosophers like Chalmers (Chalmers 2011, 2012) actually concede something like a) but argue that the gap between a) and c), or between a) and material processes in general (whether natural or artificial), can be bridged. Our characterization of pure formal structures does not *exclude* the possibility of materialization. One has to show this. So our characterization of pure structures is not question begging. Our aim will be to show that pure structures cannot be materially realized *in any determinate way*, but only imperfectly and approximately.

[46] See (Oderberg, 2007) for an example of "virtual containment". A formal structure is not what it is and does not differ from another structure, whether formal or material, *in virtue* of some physical property or of some property that obtains on account of some physical property (e.g., spatio-temporal location, genetic makeup, neurological processes, and what have you). Of course, a mathematical Turing or von Neumann machine is structurally different from a physical Turing or von Neumann machine not in virtue of *some* physical property had by the latter or *some* physical property had by the former but in virtue of the fact that the latter has a physical property at all, therefore being subject to whatever physical laws apply at the relevant level. A mathematical Turing or von Neumann machine cannot be differentiated nor individuated by the sorts of properties that could differentiate and individuate any object that falls under the category of 'physical beings'.

[47] Aristotelian forms, i.e. essences, are neither pure formal structures nor mixed formal structures. In general, Aristotelian forms differ from both pure formal structures and mixed formal structures in that they are (1) not complete *in se* but rather constitutive of something else, (2) have neither identity, existence, nor completeness apart from the things they are constitutive of, and (3) completeness, identity, and existence in their case requires materialization and a *specific*

mixed pure structures. (Though I have more to say about this in Chapter 4, nothing hinges on it).

c) *natural or material structures*. Like benzene rings, iron, and cells, these are structures whose existence, identity, and completeness are constituted by, and require and are dependent upon, a *specific* kind of material, and thus require not merely some materialization but a specific kind of materialization. (The issues of artificial structures and synthetically engineered structures will not be discussed.)

The second category, i.e. b), includes a plethora of structures which admit of no single, univocal ('one size fits all') definition. From musical pieces like Paganini's 24 Caprices for violin to bank accounts and legal documents like affidavits, this category is full of structures which can be said to be 'mixed' in different ways, some being closer in the spectrum to pure structures, others closer to material structures, depending on the degree to which their identity, existence, and completeness is dependent on some mode of materialization. The point is that nothing in the second category is classifiable as purely of the first or of the third category. The third category, i.e. c), is the category containing all structures which fall under the scope of any actual, or potential, natural science. It concerns *real* natures, that is, structures found in the natural order of things (physical, chemical, biological, neurological, geological, astronomical, etc.).[48] What is important for our purposes is the difference between structures of the first category and structures of the third, which are organized matter or energy and require an appropriate material and environment. No such thing is true of structures of the first category. This is why formal systems and rules are not subject to such things as the laws of thermodynamics or particle entanglement. There are no particles in Peano Arithmetic. It

---

kind of materialization. (A fourth condition could be added, namely, (4) they are *not* the expression and creative product of human understanding and intellectual activity—the implication being that pure formal structures and mixed formal structures are both, in different ways, ontologically dependent on human intellectual abilities and activity. But since this is not a book on the philosophy of mathematics, I will not develop this claim here nor attempt to solve the problems that it may give rise to, like how to account for synthetically engineered natural kinds, which satisfy the conjunction of (1), (2), and (3) but fail in a sense to satisfy (4).) Whether the conjunction of (1), (2), and (3) is true of *all* substantial forms, e.g. of man or of angels (if they exist), is an issue I cannot consider here. The conjunction of (1), (2), and (3), however, does capture the defining ontological features of Aristotelian forms for natural beings; if it so happens that it does not apply to, say, human beings, then humans would have to be considered the exception to the (natural) rule rather than a counterexample to it. Can functionalism avail itself of the notion of non-pure abstract functions to characterize psychological states? Maybe, but this will leave unexplained how we think in formally determinate ways. At any rate, would it succeed as an explanation of non-intellectual or non-cogitative psychological powers or states? It won't. We will see why in chapter 2.

[48] Synthetic natures or structures are an interesting but complicated case. For in a sense they very well fall under the third category, but not univocally. They require human intervention to come about, but the end result *can be* the same as what would come about by natural processes. It all depends on what kind of synthetic structure we are talking about. Take synthetic diamonds. Such synthetic structures are not strictly speaking artefacts, at least not in the way chairs, cars, beds, watches, or clothes are. Synthetic natures are *real* (e.g. synthetic diamonds are *real* diamonds, just not *natural*), but they do not come about naturally. And although they come about artificially (e.g. in a laboratory), they are not artefacts either. By artifacts I mean objects whose constituents are not intrinsically and independently ordered to the (subsistence of) the artificial structure that comes about by human intervention (e.g. a bed). But synthetic diamonds are as much diamonds as natural, non-synthetic diamonds (e.g. same chemical composition, crystal structure and physical properties). So they are not readily and univocally classifiable as artifacts. Synthetic natures come about the same way as artefacts do (that is, by the intervention and application of practically skillful knowledge), but the end result of the artificial process may be a real nature rather than an artefactual nature. But there are differences even amongst synthetic natures, for though it makes sense to say that synthetic diamonds are real but not natural diamonds it makes no sense to say that synthetic steel is real but not natural steel since there is no such thing as non-synthetic steel (i.e. there is no steel in nature). Steel is a purely synthetic structure, and so are plastic, paints, medicines, concrete, and many more. Be that as it may, the concept of 'artifact' should not be taken as applicable only to structures of kind b). For a) and b) could both be, in analogous ways, artefactual. I leave that open.

makes no sense. Each category belongs to a different ontological and conceptual order. The question is whether the 'gap' can be bridged, as in wholly physical systems capable of computing and implementing structures of the first category.[49] This is the crucial question. The first two chapters answer this in the negative.

These distinctions will recur throughout the whole monograph, and we will elaborate each at different points but only to the extent that it is necessary to do so. A work on Natural Philosophy will have to await another occasion.

5. WHY FORMAL UNDERSTANDING?

Understanding is a ubiquitous term. I have said much about understanding already, but I feel it is important to specify why I have introduced the concept of understanding into the discussion of KRA. First, because even though it is one of the most astonishing and unique capacities of human beings to acquire understanding of pure formal structures, it has at best been generally taken to be one of the least problematic aspects for a naturalistic account of our cogitative powers (see, for example, Chalmers' allocation of formal activity to the category of the "soft" problem of consciousness), and as being readily susceptible to functionalist- or computationalist-cum-physicalist explanations. But lack of proper attention to the concept of understanding when investigating the concrete physical realization, implementation, or application of pure functions, operations, and rules (otherwise labelled 'concrete computation', cf. Chalmers 2011) has led many philosophers to ignore the constitutive role that understanding plays in formal acts and activities, and to conflate what is in reality an *activity* (the exercise of an ability, of a skill, consequent upon understanding, which is acquired and learned through training) embedded in normative practices (e.g. mathematics, logic, language acquisition) with *processes* (which are anything but the exercise of an ability, of a skill, consequent upon understanding, and are neither necessarily acquired nor learned in training).[50] We wish to show that nothing could be farther from the truth.

---

[49] Another question is how structures of the first category can be used to study, explain, model, and simulate structures of the third category. This is the case for sciences that study categories of the third (and some of the second, e.g. harmonics) with structures of first category. That is, what about 'mixed sciences', i.e. empirical sciences in which mathematics is applied? We will have little to say about this in this monograph. In addition to this, structures of the third category, though they cannot be used to study structures of the first, they can be used (with human intervention) to model or simulate structures of the first. This is physical modeling, simulation, and representation of mathematical (e.g., geometrical, topological, etc.) structures, as we see in the physical display of a Euclidean sphere projected on a computer monitor. Again, we will have little to say about this in the present work.

[50] Firstly, "process" is spoken of in many ways, to be sure. One can talk about the process of learning, about legal and political processes, about the application and admission process at a university, and so forth, without the risk of confusing that with material processes. That is because in ordinary discourse, there is simply no interest or concern with any physicalist reductions. Also, physical process may correlate and underpin activities. Definite and complex processes in the brain, for example, take place when a person begins to learn a new language. Yet neither the learning nor the acquisition of the linguistic abilities can be reduced to such processes, which may nevertheless be necessary for the acquisition of the relevant abilities. Secondly, it might seem that I am begging the question against the transformational-generative grammarian (T-G grammarian, for short) by insisting that the ability to carry out formal operations and follow pure formal rules is learned and acquired. For, the T-G grammarian would say, not all pure formal structures are created equal and though such an observation might be true of all non-linguistic formal structures, it is not true of all linguistic ones—by which she means the structural rules or principles and parameters of Universal Grammar (UG, for short) which are innate. Likewise, by insisting that in order to *v* one must understand *v* I seem to be begging the question against the T-G grammarian who holds that some pure formal structures (namely, those proper to UG) are, though carried out and applied, not understood. They are "cognized", which is an unconscious computational process. According to the T-G grammarian, we come with the built-in *ability* (not merely *capacity*) to carry out certain formal operations, of which, coincidently, we need understand nothing about. Therefore, if *v* is a formal operation that belongs to UG, then (1) we are *able* to carry out *v* or to engage in *V*-ing without understanding anything of *v* (for, we *v*

The second reason is due to Kripke himself, who developed his skeptical paradox (Kripke, 1982, 1981) by means of mathematical functions and operations, as opposed to mere color predicates (Goodman) or the translation of words and sentences from one language to another (Quine).[51] As I've pointed out already, to follow and apply a formal rule is not merely to think or talk *about* something, even if competently, but to proceed and engage in thought in the formally determinate ways of which the rule is dispositive. Here we find a crucial difference between Goodman's 'green/grue' considerations (Goodman, 1983; cf. Hacking, 1993) and Quine's 'gavagai' example, on the one hand, and Kripke's 'addition/quaddition' considerations, on the other. We do not green or grue anything. There's no such thing as 'greening' (in the sense in which 'green' is ordinarily used, to mean the color green) or 'grueing', except in the equivocal sense of painting something green or painting something grue (which one?). But we do add or quadd, and there *is* such a thing as adding and quadding. 'To add' signifies to carry out a formal operation in accord with some formal rule or function which guides and informs the activity in determinate ways, and which must be understood in order to be carried out or applied. (It is our claim that formal rules cannot be followed ignorantly and accidentally.)[52] This is true not merely of functions like addition but of algorithms for solving functions, like 'to borrow and to carry'. One cannot do these things if one does not understand them (that is, have a grasp as to what it is to do them). But once one understands what it is 'to borrow and to carry', one is able to borrow and to carry (not infallibly, of course; understanding does not rule out the possibility of mistakes). 'To follow and apply a formal rule', 'to compute a function', 'to carry out a formal operation', and the like, are all descriptions of something we (can) *do*, not merely of something we can think or talk *about*.

Quine's indeterminacy of translation (Quine, 1970, 1969, 1960) is closer to Kripke's argument than Goodman's (but see Kripke, 1982, p. 58 for a somewhat different assessment), but even Quine's indeterminacy argument did not put into question that we could engage in translations. Furthermore, for Quine, the set of admissible determinants for the determination of definite meanings was taken to be purely behavioristic and therefore the premises for the indeterminacy of translation argument are precisely behavioristic.[53] The question was, rather, whether any possible

---

unconsciously) and (2) the ability to *v* is neither acquired nor learned but is, rather, innate. Though the T-G grammarian's position seems to me to amount to sheer nonsense, the accusation of a *petitio* is simply not to the point. For even if the TG-grammarian's position were correct, in which case talk of 'formal understanding' would be on the whole correct but without use when considering the pure formal operations and rules of UG, there would still be a formal *activity* and an *ability* to be accounted for. And if its formal, it will be subject to our argument: *V*-ing cannot be wholly a physical process, and therefore neither can the ability to *v*. To be sure, this refutes the generative grammarian's (or, rather, Chomsky's) neurobiologism concerning the constitutive and genetic solutions to UG, but it does not refute UG as such. For a fuller treatment of KRA and Chomsky's generative approach to grammar, see chapter 6.

[51] In fairness, Kripke takes the skeptical paradox to be completely general, not limited to mathematical problems but applicable to *any* rule or word of language including "the language of color impressions." (Kripke, 1982, pp. 58–60) In our exposition of the skeptical paradox, and of Kripke's argument against computationalism, we make no such generalization. That it is applicable to mathematical and logical problems, we show. That it is applicable to language as a whole, we don't. We leave that open, except when it comes to theories that take language to be, wholly or partly, computational or computationally generated. Again, see chapter 6 for an example.

[52] This much should be clear from the preceding discussion. The idea is that a necessary condition for there being a difference between behavior that merely conforms and accords with a rule and behavior that actually follows or applies the rule is the element of understanding. This is true of rules in general. A driver who consistently stops at a red light because he wants to admire the beauty of the color red or because his legs cramp at the sight of a red light is not following the relevant traffic rule; his behavior is one of mere conformity to the rule, not of following and applying the rule.

[53] This is tricky, because if the consequence of the indeterminacy of translation is that there cannot be any such thing as meaning, then there surely cannot be anything as translating unless translating is a formal activity without determinate content (much like rules for manipulating symbols that do not have any semantics assigned to them). But then it is hard to see what it is meant by *translating*. At any rate, if there's no such thing as translating something, i.e. no such thing as

linguistic fact (behaviorally understood) accessible to the radical translator[54] could determine which of the possible incompossible translations of the same set of words and sentences of a foreign language was correct. The point wasn't that it was indeterminate whether a translation had taken place in the first place, but rather that it was indeterminate whether the translation was correct (or incorrect). Quine took this to mean that there is no fact of the matter as to what the sentences of that language mean, since all the linguistic facts accord with an indefinite number of incompossible translations. Kripke's skeptical paradox puts into question addition *as an activity*. It is not merely that there's not fact of the matter about whether we *mean* addition or quaddition when we use the word 'addition' or 'quaddition' or the symbol '+', but also that there's no fact of the matter about what we are *doing* when we carry out operations with the symbol '+' (or '⊕'). Are we *adding* or *quadding*? If there is no fact of the matter as to which formal rule we are following, then there is no fact of the matter as to whether we are adding (or quadding).

Thirdly, to paraphrase Hacker (see e.g. Hacker and Baker, 2014, p. 40ff)(see Hacker, 2008, p. 40), to understand a formal operation or formal rule is to know what that operation or rule is; but if one knows what a given formal operation or formal rule is, one knows what it is to carry it out or follow it, and if one knows what it is to carry it out or follow it then one is in a position (ability-wise) to carry it out or follow it.[55] By coming to understand a pure formal operation, rule, function, I therefore mean the onset of a definite set of *abilities* with respect to what has been understood and which can be manifested in a variety of ways and in varying degrees of complexity, breath, and depth depending on the degree of understanding. Ways of manifesting and exemplifying understanding are e.g., defining variously what has been claimed to be understood, providing explanations of the rules, reformulating it for heuristic purposes, carrying it out and exemplifying it, clearing out confusions or pointing out misapplications, etc. Of course, one can understand a formal operation without manifesting it and one does not need to manifest it in order to have the ability to carry out the relevant operation; but this is not to the point. We are not behaviorists. The point is that if one has understanding, one *can* manifest it (that is, have the ability to engage in the sort of activity that manifests the possession of understanding).

Earlier, I made a conceptual distinction between the *semantic* or *intentional* aspects of understanding and the *formal* aspects of understanding. The semantic or intentional aspect is what the thinking is *about*—otherwise called *content*, "the variable component". The formal aspect is what structures the thinking, what determines the *kind* of intellectual act it is (i.e. the formal aspect is the *structure*, "the invariable component"). The distinction should make it obvious that, in practice, one can have one aspect without the other when exercising one's understanding of a formal operation. One can think about a certain formal operation without thereby carrying out that formal operation,

---

translation, the argument for the indeterminacy of translation cannot even be made. For some notable discussion about the indeterminacy of translation argument, some in comparison with Kripke's, see (Foellesdal, 1973; Evans, 1975; Ebbs, 2001; Soames, 1999, 1998a; Wright and Wright, 2017; Miller and Wright, 2002; Hylton, 2018; Morris, 2006; Miller, 2007; Shapiro, 2007).

[54] The radical translator is conceived in an *idealized epistemic state*, such that all relevant linguistic facts are accessible to him. The linguistic facts will all be of the same kind, namely behavioral, since these were the only kind of facts that Quine took to be available in translating. This notion of an idealized epistemic state is important, and it is likewise implicit in Kripke's argument (cf. Kripke's appeal to an omniscient being 1982). For it enables both of them to make a *constitutive,* rather than a merely *epistemic,* case. More stuff of the same *kind* won't change the outcome, for the problem does not lie in the quantity and complexity of facts considered, but on the *kind* of facts considered. And if it's a matter of *kind*, considering ever more or even all the members of that kind will not help. When we come to the Stabler-Lande formulation of Kripke's argument, we will see that the idealized epistemic state has been actualized by the simplicity of the physical device which Stabler constructs.

[55] Note the difference between "know *what v* is" and "know *that v* is". A person can know that there is such a thing as commutation or derivatives but not know what that is.

and vice versa. Though I do not think that one can be said to *understand* a formal operation[56] without knowing what it is to carry it out and thus without being able to carry it out, to understand a formal rule without knowing what it is to follow it and without being able to follow it (indeed, the ability to genuinely follow a rule requires the ability to think about it and grasp it and the ability to think about it and grasp it requires the ability to genuinely follow it)[57] one can, of course, think about addition without adding (intentional aspect) or add without thinking about addition (formal aspect). One cannot add addition. That makes no sense. The fact that one understands what addition is does not mean that one is therefore adding at every moment, nor that one can add without ever making mistakes.[58] The fact that one understands what addition is just betokens, in part, that one *can* add and that one knows what it is to add and is thereby able to grasp and spot incorrect applications of the addition function (constrained by the degree of understanding), something which is consequent upon understanding what addition is. Understanding does not entail infallibility.

In short, if one understands what addition is, then one knows what it is to add and has the ability to add (that is, *genuinely* add, the exercise of an ability with a grasp as to what is done, not by route or imitation the way a parrot is said to "speak").[59] But one cannot be said to *understand* addition if one

---

[56] To have a vague idea of addition is not to understand addition. On the contrary, it's precisely *not to*. One can also *mis*understand a formal expression (e.g., taking x^n to be equivalent to $n \bullet x$ and thus assigning e.g. 2^3 and 3^2 the exact same value, "6"). On a slightly different note, I do not think these considerations are extendable to understanding in general. One cannot understand a formal operation without being able to act in accord with it. One comes to understand a formal operation by acquiring the ability to do it. This is not the case with swimming, metabolism, atom decay, investments, etc. One can know what it is to swim (to a certain extent) without having acquired the ability to swim. One can just point to another person swimming. One can understand metabolism without having the acquired ability to metabolize (of course, this makes no sense, for metabolism is not an acquired cognitive ability). But not so for formal operations and rules, which constitute ways of thinking, deducing, inferring, calculating, etc. One can of course c*laim* to understand what addition is without knowing how to add. But pointing to another person solving a differential equation on a black board will not do because you would not know (would not be able to know) whether the person was really solving a differential equation or doing anything meaningful at all.

[57] Manifesting behavior that merely conforms with the rule but is not an instance of following the rule (e.g., when a student answers "4" to the problem "2^2 = ?" because "4" is her favorite number or because it was the numeral written down for '2^2' on a cheat code). One is in a position (i.e., one is able) to follow a formal rule only when one understands the formal rule. Though this is more or less true of all rules, it is strictly true of formal rules. Take Sudoku. A quadriplegic may understand the formal rules of Sudoku, even if she cannot *manifest* that understanding herself in the typical way—except by some external aid to write down the answers (assuming she is playing the game on a Sudoku paperback). But this seems to be a problem not of a lack of ability with respect to the game but of the *manifestation* of that ability which is effectively blocked by accidental circumstances (e.g., neurological impediments). Of course, there is a lack of some ability (say, with respect to the motor movements necessary for writing) *but not with respect to the formal rules of the game*. You cannot claim to understand the formal operation of conjunction and have not the slightest clue as to how to conjoin and what it is to conjoin. You cannot be said to understand what it is to simplify a radical and yet not be in a position (not have the ability) to simplify a radical. It is true that one can have an understanding of addition over natural numbers without knowing what it is to add fractions. But this is an issue of degrees of understanding, not of understanding as such.

[58] To what, if anything, are mistakes to be attributed? We consider this briefly in chapters 2 and 3.

[59] If one understands the DeMorgan rules of replacement, one knows what it is to apply them and has the ability to do so. This does not mean that one will never be misapply DeMorgan's rules. Mistakes and misunderstandings occur. Likewise, none of this implies that understanding a pure formal structure implies having the ability to engage in *any* possible kind of formal activity with respect to that formal structure. One can understand Goldbach's conjecture and not be able to prove it (indeed, at the moment of writing, the is no known proof of the conjecture). But if one understands it, one must be able to at least explain it or have (e.g., manifest) some sort of competence with respect to it. One can understand the square root function and thus be able to compute it without having any understanding as to its graph and representation in analytic geometry. When someone is taught and understands some mathematical operation over positive integers, for example, but has not been taught how to carry out the operation over negative integers, she might not know what to do and might misapply the rules for the operation when confronted with the requirement of an application that lies outside her understanding (say, a problem with negative integers); nevertheless, she does have the

does not grasp the addition rule and doesn't have the ability *follow* the addition rule (in other words, doesn't know how to add). The argument to be defended in the following pages implies that the physical cannot operate *in terms of* pure functions and therefore cannot *determinately* follow pure formal rules and carry out formal operations. The physical, at whichever level, cannot be structured in any such formally determinate way. Now, understanding need not be complete even if the pure formal structures themselves are *complete*. One can understand addition without understanding all its uses, extensions, and applications in different domains. If, however, one really does understand addition, then that means that one knows what it is to add, even if not completely (e.g. as when a child has not yet been taught that there's addition over negative integers) and one can manifest to whatever degree that very understanding by adding, or explaining how it is done, or illustrating a case of addition, or pointing out incorrect applications, etc.[60]

If the way in which I have couched some aspects of the concept of understanding causes confusion or revulsion, the reader is free to excise it completely from her reading. The point I want to bring home is that the ability (and consequent exercise of that ability, i.e. the activity) to think in terms of, to carry out, or to follow pure formal operations and rules cannot be a purely physical phenomenon. Our thinking can be structured in ways that are categorially different from the ways in which anything physical can be structured. Formal understanding begets the ability to think in ways

---

ability to carry out the mathematical operation in question but only to the extent of what she understands, that is, the operation as defined for positive integers. Understanding comes in degrees and can be compartmentalized. This is what I mean by saying that coming to understand some formal structure is, at least partly, the acquisition of a set of abilities with respect to that structure. If it's an equation, or theorem, or axiom, what has been understood, then one is able to do certain things with respect to such items whether it be define, explicate, exemplify by proof, reformulate, simplify, etc. If it's a formal rule *f* what has been understood, then the ability (or set of abilities) acquired is that for following the rule *f*. But one cannot have a ability for *f*-ing if one does not understand what is *f* and what it is to *f*. Does the converse (i.e. "if one is able to engage in formal activity (with respect to) *f*, one understands the formal structure *f*") hold? 'If one knows what it is to add and thus has the ability to add, one understands what addition is.' I think so, but it will all depend on what we mean by 'understand'. Naturally, "understanding" is spoken of in many ways. In one sense, one can be said to understand a formal rule when one has grasped how the rule works and what one must do to follow it. In another sense, one is said to understand the formal rule when one grasps *why* it is to be followed and *why* it works. I am using "understand" in either way. For instance, take the 'estimate and check' method for calculating the square root of a number. One can master this technique without knowing exactly why it works, i.e. why it is so that when carried out correctly it produces the square root of a number. But that is still a case of formal understanding, as the technique is a formal rule (or, technically, an algorithm) for calculating a square root. The student doesn't even need to know that such a technique yields the square root of a number. But she does need to understand the formal rule, or algorithm, to be able to follow it. For to understand a formal rule or operation is, partly, to understand what it is to apply it or carry it out. If the student doesn't understand what it is to carry out the "estimate and check" algorithm, she cannot to know what it is to follow it and will not be able to carry it out. Most people can add and do understand what addition is, but they be hard pressed to provide a definition of addition, let alone a set-theoretic proof. Most people would simply manifest their understanding by just adding and showing others how it is *done*. A deeper understanding, at the level of the mathematician, gives one a wider set of abilities to deal with the arithmetical function. The analogy with how understanding a language works might be useful, and it is more to the point than the "swimming" analogy. If one is able to grasp and think in terms of a language, then one understands that language, and if one understands a language, one is able to grasp and think in terms of a language to whichever level of competency one has.

[60] We agree with (Stokhof and van Lambalgen, 2011a, 2011b; Stokhof and Lambalgen van, 2017) in that individualism is untenable. Even though our argument (KRA) focuses on the capacity and the ability to do mathematics and logic, which are part and parcel of individual human beings, we do not clam that the ability and activity are to be explained by features of the individual alone. We reject individualism but not the individual in the explanation of formal understanding and activity. The capacity might be susceptible to species-specific analysis, but the determination of the capacity in the acquisition of formal understanding and its exercise in actually carrying out formal operations and applying formal rules is, explanatorily, practice-bound. We think something similar is true of the capacity for language (what is linguistically determinable in the individual) and of linguistic understanding (what is linguistically determinate in the individual who has mastered a language).

that are determinate in formal structure, and nothing in the physical can *on account of its physicality alone* be determinate in formal structure. Though it seems to me that to understand a pure function, a formal operation, or a formal rule, signals amongst other things the ability to compute it or follow it, I am more than willing if push comes to shove to renounce any talk of understanding for the sake of the argument to be put forward. What must not be overlooked is that we are investigating (in epistemic order) an activity, an ability, and a capacity. Whether the acquisition of the ability is tantamount to the acquisition of understanding is, though crucial, something which can be postponed for later debate. But that the ability in question is an ability whose exercise constitutes an activity that cannot be naturalized mustn't be lost from sight, as it is the fundamental claim of this monograph. This is a dialectical maneuver to allow whoever objects to the concept of understanding as employed here to not let quibbles over the concept of understanding obscure the nature of the argument at hand. But enough of dialectical concessions. Since understanding is (at least in part) ability-like and formal understanding is (at least in part) an array of abilities for carrying out definite formal operations, it follows that anything which cannot carry out and follow formal operations and rules cannot be a possible subject of formal understanding. Even if the abilities in question are not the whole of the understanding, they are at least a constitutive component of it.

6. PRACTICE-BOUND UDERSTANDING: PRACTICES, TRAINING, RULES

Let me finally say something about the nature of formal understanding by drawing a couple of analogies with non-formal normative practices in the hope of illustrating that there is no intelligibility to the concept of formal understanding outside, or independently of, the practices which make it possible and in which it plays a constitutive role. Formal understanding is acquired through experience, practice, and training. There is such a thing as 'staying in mathematical shape', as any professional mathematician would tell you, and there is such a thing as a normative practice (or set of practices) called 'mathematics' or 'mathematical activity'.[61] Different practices, even different formal practices (e.g., topology and computability theory), determine what counts as having understood the relevant subject. There is no "one-size fits all" analysis of 'understanding' because what counts as 'understanding' is sensitive to the practice in which understanding is made possible. Just consider the difference in the manifold uses of 'understanding' in, for example, law practice (e.g., procedural law), economics (e.g., markets), sports (e.g., Olympic weightlifting), games (e.g., chess), and mathematics (e.g., number theory). Understanding also comes in degrees with different layers of complexity, like the fact that both algebra and geometry can be understood by someone who has yet no clue as to the fact that one can use the former to model the latter (analytic or Cartesian geometry).

Consider the following example, which depends both on insight and on prior understanding of algebra. Suppose a pupil is asked to indicate in sigma notation the following series:

$$1)\ 1 + 2 + 4 + 8 + \ldots + 64.$$

Understanding of certain formal operations is necessary for one to able to *see* the abstract pattern or structure in question (that is, have insight into 'the formal skeleton' of the series, as Lonergan would say):

$$1')\ 1 + a + a^2 + a^3 + \ldots + a^n.$$

[61] See (Hacking, 2014) for the variety of definitions of mathematics and the difficulties of giving a one-size-fits all definition of 'mathematics' that is satisfied by all mathematical branches or fields.

But one cannot have insight into the structure of the series unless one has a prior understanding of the operations and rules which each series exemplifies. The structure would still be there, wholly intelligible, but the relevant conceptual lens which would allow the student to discern and hit upon it has not yet been acquired. But suppose the pupil does have understanding of algebra and the ability to see 1´) in 1) and to state it thus. Even then, if she has no idea what sigma notation is, she would not understand what is being required or asked of her. For she does not understand what it is to indicate something in sigma notation, even though she might be able to see and state the formal skeleton of the series in non-sigma notation. So, the student might be able to go from 1) to 1´) without being able to go from 1 to

$$1^*)\ 1 + 2 + 4 + 8 + \ldots + 64 = \sum_{k=0}^{5} 2^k$$

and from 2, and/or 1*), to the more general

$$1^{*\prime})\ 1 + a + a^2 + a^3 + \ldots + a^n = \sum_{k=0}^{n} a^k$$

The pupil does not know what to do nor what is being required of her because she does not know how, in the relevant practice, the symbols that conventionally constitute sigma notation are *used*. And this is so for all mathematical (and logical) notation. She needs to be introduced to this, that is, taught what it means to write a series in sigma notation, what the given symbols mean and how they are used in the relevant domain of mathematics. Without this, she cannot even tell whether these apparent symbols are indeed symbols and therefore mathematically meaningful—or meaningful *tout court*. Only after such instruction or introduction can she begin to understand and be in a better position to have insight (if at all) into the formal skeleton of 1) as given by sigma notation. Note that when the pupil is asked to explain or define "sigma notation" (say, in a midterm exam) and to indicate in sigma notation the series in question, she is generally being asked to *manifest* her understanding of sigma notation, that is, her grasp and dominion of the relevant technique and operation (i.e. what sigma notation is and what it is to indicate a series in sigma notation).

All of this should make it clear that understanding comes in degrees or levels or kinds. This is the exact same point that Lonergan made when he coined the expression "higher viewpoints", a rather obvious but neglected feature of any learning process embedded in a practice where expertise or competence is possible (thus, where teaching, training, and mastery take place). In general, when it comes to skillful practices like mathematics, ever higher or deeper understanding always builds on prior, lower understanding, insight upon prior insight, there being things that cannot be understood unless certain other things have been understood first. One cannot hope to understand non-Euclidean geometries without any understanding of Euclidean geometry, or calculus without any understanding of algebra. Let me make an analogy. When we come to non-formal practices, like strength training for instance, the principle is much the same: there is much that cannot be understood unless certain other things are understood first.[62] One cannot be said to understand the

[62] Strength training is the basis of both weightlifting and powerlifting, and indeed of any training program that aims at generating and improving *strength*, even up to the point of individual maximal strength as is the case in Olympic weightlifting and in powerlifting competitions. There are different programs and systems for strength training. (The

power clean (with a front squat) without having any understanding of the squat, in particular the front squat. The notion of understanding at play here is *analogous* to the one at play in mathematics, not univocal to it, since 'understanding' in strength training is intrinsically tied to *embodied movement patterns*: one thing is to read about the proper form or mechanics of a squat, whether front bar, high bar or low bar, but another thing is to be able to execute it with proper form and mechanics, to apply the theory and grasp how it should *feel* and look *in you*, and to be able to *see* it manifested (or misapplied) in a variety of exemplars and occasions. This is the "aha!" moment in training, the *embodied* insight into the proper or improper form of a pattern of movement under considerable weight load; and even this (the embodied insight) comes in degrees and with the possibility of faulty judgements.

A competent strength training coach (e.g., an expert in the practice and community of weightlifting) will able to teach the way of training properly, provide effective advice, tips, and tricks, will be able to explain the *raison de etre* of their effectiveness, will know how the relevant exercises (e.g. a low bar squat) should look relative to the individual (taking into account differences in body types, in femur and torso length, in shoulder width, etc.), will be able to spot errors and correct the potential sources of discomfort (so long as these are caused by improper form) and competently suggest training programs and modifications to the program where necessary, etc. Likewise, strength training is a *normative* practice: there are wrong and correct ways of training, misunderstandings and mistakes, proper and improper form, rules as to what counts as failing a rep or completing a rep, standards as to what counts as excellent and what counts as average and so forth. In short: 'understanding', a bit like 'truth', is a *plastic* notion whose full or adequate specification depends on the practices, and contexts in which it is embedded and acquired. Strength training and mathematical training are both practices aimed at acquiring a skill and mastering a set of techniques, at improving on such skills, and perhaps at contributing to the advancement of the field. They both have performance tests at every level (e.g., high school calculus exams) that test a *skill*, an *ability* to do certain things, and make way for competitive events (e.g., William Lowell Putnam Mathematical Competition). Mathematics *qua* activity is like strength training or Olympic weightlifting a skill-based normative practice. Yet each one trains wholly (i.e. categorially) different aspects of the human person and thus require a (categorially) different set of abilities. Strength training is about training and strengthening one's *body* in a methodic and systematic way and acquiring the set of abilities that will allow you to effectively produce more force against measurable external resistance. Mathematics and logic are paradigmatically intellectual endeavors and as such train the intellect for certain purposes. Even linguistic (with qualifications) understanding can be said to be the outcome of a categorially different set of 'training' or 'stimuli', though no goal seems to be set in the case of learning one's mother tongue and no academic instruction seems necessary (though language learning in adults is normally more akin to training as discussed in the case of strength training). At any rate, mathematical and logical understanding (and linguistic understanding, with some qualifications) are *analogous to the strength* acquired through systematic and effective training in a given normative, established practice (in this case, barbell training) and *to the abilities* that enable the manifestation of that strength and the development of further (feats of) strength.[63] Yet the proper

---

difference between weightlifting and powerlifting is that the former uses overhead movements, such as the snatch and the jerk, whereas the latter uses movements that are not directed vertically overhead, such as the squat, the bench press, and the deadlift. There is no such thing as 'Olympic powerlifting', but there is such a thing as Olympic weightlifting.)

[63] Though it is well known that 'genetics' or the natural, inborn physical constitution of an *individual* makes a difference in sports (e.g., in natural and potential strength) it is hard to say whether something is similarly the case with respect to intellectual capabilities. It is not easy to say what accounts for inborn intellectual 'gifts' if such things exist. But it is true that some individuals seem naturally more gifted than others in certain disciplines, including mathematics and logic, but systematic training and practice are still necessary for that natural potential to be actualized and to be taken to the point

manifestation of understanding, and even the criteria for understanding will differ in each case—though in each practice understanding will be, at least in part, ability-like—simply because they are different *kinds* of practices and activities. This should all be taken as pointing out important analogies between these two kinds of activities, and not as saying that there aren't important disanalogies between them. For there are (e.g. you can have natural innate strength, but you cannot have natural innate formal understanding *pace* Chomsky; mathematical objects are at least *sempiternal* whereas the barbell, the weight plates, and overall lifting equipment are not, and so forth). The point is that mathematics is, like many other practices, *a human activity* situated in normative practices with standards of correctness and excellence, subject to the inevitable aspects that any human activity embedded in normative practice where excellence is possible is subject to. Like any another human practice, there is the tripod of a definite set of capacities (condition for the possibility of abilities), abilities (acquired by training, mimesis, practice, etc.), and activities (abilities exercised in accord with

---

where excellence and feats of skill can happen. Those who are naturally gifted will most likely advance way faster than the average Joe (if the appropriate social, cultural, psychological, physiological, moral, and intellectual conditions for the development of the relevant skills are present and held somewhat constant) and will be able to achieve greater and seemingly impossible feats of skill in their respective fields (e.g. Mozart in music, Usain Bolt in track and field, Srinivasa Ramanujan in mathematics, etc.). Again, this is a further case of analogy between non-formal practices like sports and formal practices like mathematics, logic, formal linguistics, etc. It is only an analogy, and there are therefore important disanalogies that we mustn't ignore (see, for example, our treatment in section 6). In the case of language, the analogy partly breaks down when we come to children's acquisition of their first language(s) and the corresponding linguistic understanding (though not when we come to second or non-native language acquisition in adults). For, the acquisition of a first language in children seems quite spontaneous and not the result of any purposeful training, at least not in the way in which chimpanzees are trained to communicate with humans or adults are taught a second language. Some call the capacity for language that children have a 'second-order ability' (Hacker and Baker, 2014, p. 158), since children seem able to acquire with relative ease and short amount of time a new language, a skill which seem to fade or begins to fade at around the age of 10. Whether the use of 'ability' here is adequate is debatable. It just seems that talk of capacity, grounded in our rational nature and truths about our physiological development, is enough to distinguish a child's flexible capacity for language acquisition in comparison with that of an adult's. Capacities, precisely because they are natural, are subject to all of the effects of developmental processes which beings like us go through. Also, how many 'second-order abilities' are there? A healthy young man from the age of 16 to 35 can be said to have a second-order ability to get extremely strong and build incredible amounts of muscle mass, an ability which fades or begins to fade after the age of 35 and is almost impossible past the age of 50 (except through human hormone growth). Furthermore, when it comes to the technique of the different kinds of lifts (e.g. clean and jerk), it seems that the younger a man or a woman begins to be exposed to the lifts (at least to their proper technique) the easier technical mastery will be. Of course, we can call that a second-order ability but the same could be said of a child's more flexible capacity for acquiring say musical (e.g., violin), dancing (e.g., ballet), mathematical (e.g., algebra), and sport (e.g., basketball, boxing) skills. The point is, *why* should we call it a second-order ability at all? I think we should reserve the concept of 'capacity' for that which we are naturally endowed with, admitting developmental effects and variations in the capacity, and the concept of 'ability' to skills which are acquired. Back to disanalogies. Unlike mastery of mathematical and logical rules, excepting severe cases of genetic and neurological anomalies, every child acquires mastery of at least one language without seemingly much effort and instruction. Linguistic understanding is a much more universal and systematic phenomena in human beings than mathematical and logical understanding is. Another disanalogy: talk of 'linguistic geniuses' is usually with respect to people who can learn, whether children or adults, very many languages with relative ease ('literary genius' being reserved more to those who have a dominion of prose that is extremely rare and exceptionally good within a given language) not to people who master a language. When it comes to mastering a language (i.e. a mother tongue) it seems all healthy human beings do relatively equally well here, which is probably why being a language speaker (that is, having a mother tongue) is not generally considered a feat of skill, a sign of genius, or praiseworthy. Yet the disanalogies should not be overblown. Language is nevertheless taught, learned, and mastered and this requires, like any other sort of activity in which mastery is possible, a normative practice and community (e.g. the linguistic activity of the community to which the child belongs), with standards of correctness and paradigmatic examples for the application of the rules of the language and a complicated web of corrective measures for misapplications. None of this needs to be codified. It is all fully present in and determined to a great extent by the lived-life or forms of life of the linguistic community, the culture, etc. (Braine, 2014; Hacker, 2017; Hacker and Baker, 2014, 2005)

understood standards or effective procedures) that is satisfied and is required for any being to partake in the relevant practice.

Possession of the ability is not necessarily enough for the manifestation of the ability. For example, I might have the ability to execute a 170kg snatch but I cannot do so in the absence of a barbell, the appropriate weight plates, the platform, etc., just as one cannot exercise and manifest one's ability to play tennis (that is, one cannot play tennis) in the absence of a tennis ball, a racket, etc.[64] In the case of logical and mathematical understanding, conditions of opportunity will be different of course (e.g., there is such a thing as mental arithmetic and adding mentally but not such a thing as mental weightlifting or a mental snatch, except as an imagining). And one cannot acquire the ability to do these things without the environment and practices that constitute these activities. Something akin to this is true of an individual's logical and mathematical understanding which is acquired and made possible in part by logical and mathematical practices that constitute the relevant sorts of activities *as* logical and mathematical and in which the individual is taught and trained or in which the individual acquires the ability to do these things.[65] To be able to indicate a sum using sigma notation is an example of precisely the practice-bound feature of mathematical understanding.

---

[64] There are further distinctions to be made here that we do not have the occasion to address here for reasons of space and scope. But consider the following explanation of the distinction between an ability and the exercise of an ability. "An ability is distinct from its exercise. A person may have the ability to play tennis, but cannot (exercise the ability to) play unless he has a racket, ball, tennis court and partner, i.e. unless he has an opportunity. An opportunity is a condition for the exercise of an ability. Among opportunity conditions one may distinguish logical opportunity conditions from agent-relative opportunity conditions. No one can kick a goal without a ball, but what constitutes an opportunity for a skillful player may not be one for a beginner. One's exercise of an ability may be restricted not merely by lack of opportunity conditions, but also by 'external' restrictions (e.g., one cannot play tennis if one's legs are shackled) and by 'internal' one's (e.g., one cannot play tennis when very drunk). The boundary line between 'external' and 'internal' restrictions or constraints upon the exercise of one's abilities is not sharp, but neither imply loss of ability. Neither being shackled nor being drunk (or ill) imply loss of the ability to play tennis, but only that it cannot be exercised in these circumstances. 'External' constraints are not limitations upon one's ability, but upon its exercise. However, one might say that 'internal' constraints are. The ability to play tennis cannot be exercised when very drunk, but then it does not include or subsume the ability to play-tennis-when-drunk. When drunk one does not lose one's (generic) ability to play tennis, but even when sober one does not have the ability to play-tennis-when-drunk. The circumstances of the agent may restrict his opportunities, the condition of the agent constrains his abilities. Again, the boundary between circumstances and condition of the agent is not absolute." (Hacker, 1990, p. 131) Much more needs to be said, but we agree with the thrust of this statement.

[65] For a similar view along these lines, see (Hacking, 2014). Again, that mathematics and logic are human practices does not in and of itself gainsay the claim that mathematics and logic are concerned with (and even discover) objective truths. This might be so. Or it might not. It remains that we do mathematics, that doing mathematics is an activity we (can) engage in, that mathematics has developed through the ages, and that the same pattern of novice-apprentice-journeyman-master/expert that is found in all practices is found in mathematics (and logic). Most people do not get to the level of expert or master, only some do; others are gifted and can jump the path from novice to expert at an astonishing fast rate (consider, for example, the case of Euler in mathematics and Mozart in music), whilst most of us are stuck in between; still others though naturally gifted are either never exposed to the appropriate environment to would allow for and foster the development of their natural potential or are exposed too little too late to it and forced in the meantime to employ their genius with whatever little it is they have received (e.g., the Indian mathematician Ramanujan). What determines whether one has become a master or expert in a given field of mathematics or logic? What determines whether one is a genius in such fields? *What one can do* in the relevant field or domain. The standards of excellence and norms of correctness that are proper to a given practice (particularly a well-established and developed practice) together with its past and present accomplishments determine what is excellent and what is correct in that practice; if one can meet those standards and apply those norms, particularly with an ease and skill that makes it possible to accomplish something that is extremely difficult and/or that has yet to be accomplished, thus setting new standards of excellence in the relevant practice, one is an expert or a genius in the relevant field. This is true of musical excellence, of 'greatness' in sports, of brilliance in theoretical physics, of 'genius' in mathematics and logic, and of any practice where mastery, competence, and excellence are possible. For a more thorough development of this line of thought in regard to musical excellence, see (Ross, 1993).

This is why in our opinion there cannot be an individualistic explanatory framework of mathematical understanding if the explanatory framework is to be as adequate and complete as possible. The individual who does mathematics cannot be taken in isolation in the study of the nature of that sort of understanding because that sort of understanding is practice-bound and therefore requires a community (e.g. a mathematical community) with an established practice to which the individual is introduced and to which it owes in great part the understanding she possesses. Yet it is she who possesses that understanding once it is acquired, and our question concerns the nature of those capacities or powers of *hers* that make her a potential subject of mathematical understanding. This is what we are after. A full-blown study of the nature of mathematical understanding (which is not something we intend to carry out here) must survey what mathematical understanding consists in, how it is constituted and made possible by shared and established norms and standards of excellence that are the jewels of the mathematical community, what the precise nature of its exercise (activity) is and of the abilities whose exercise counts as doing the relevant activity, what the nature of normative practices and practice-bound understanding is, and so on. All human beings are *in principle* suitable subjects of formal understanding and we have a natural capacity for understanding, in particular linguistic understanding. But mathematical understanding seems to be a species-specific trait too, though maybe parasitic on linguistic understanding.[66] At any rate, the nature of our capacity for formal understanding is one we hope our monograph will help illuminate, even if only in part. Yet a full-blown account of the nature of mathematical understanding must consider not merely the species-specific cogitative capacities that all human persons share but also the established cogitative practices, which are not shared by all human communities but which give determination to the cogitative capacities of human persons so as to make it possible for them to acquire the abilities to engage in this or that sort of intellectual activity. It must take notice of the practice-bound feature of mathematical understanding much like an account of a tennis player's ability to play tennis competently or of a sport analyst's understanding of the game of tennis cannot do without considering the very game of tennis (techniques, standards of excellence, norms, etc.) which 'transcends' the individual. There is no mathematical understanding without mathematics, just as there is no understanding of tennis (or any other sport) without tennis (or sport). Something similar can be said of linguistic understanding, though things a more nuanced and complicated in the case of natural language.

I have hinted at the claim that mastery of a natural language seems to be a necessary basis for logical and mathematical understanding, perhaps a condition for the possibility. Depending on the natural language, the language will be either an enabler or a limiting factor in the acquisition of the more technical logical and mathematical abilities that go well-beyond what is contained in the acquisition of the relevant natural language. No one is a mathematician or a logician by simply learning a natural language, though one can (depending on the natural language(s) that figures as the mother tongue, of the culture and linguistic community, of the individual, etc.) think mathematically and logically nevertheless on account of the logico-grammatical structure of languages and to the extent that the relevant linguistic practices of the community already contain (say, by the development of certain logical and mathematical practices within the relevant community and culture) logical rules that allow for an array of inferences as well as mathematical (or proto-mathematical?)[67] concepts and rules that allow to different degrees for such things a quantification,

[66] Our argument will have as a consequence that any being capable of formal understanding has a capacity that cannot be accounted for in material terms alone. So if there are other beings (say, other species in other galaxies) that have the capacity for formal understanding, then they have a capacity that cannot be wholly a matter of matter.

[67] I have in mind a language like Pirahã (the natural language of the Amazonian tribe Pirahã) which purportedly lacks numeracy (that is, numerals and number-concepts) yet contains some quantitative concept-words that allow for some minimal discrimination, i.e. 'a few' vs 'a lot'. There is no such thing as 'discreet quantities' in the language of the Pirahã,

counting, measuring, and so forth. The issue is complex, and we wish to make no definitive claims on such matters. Our claim is merely the following: formal understanding consists in part in the ability to engage in a kind of activity that is practice-bound or practice-determined and requires the actualization of something (i.e. formal determination) that cannot be, constitutively and genetically, explained in material terms alone. Hence, the ability to engage in a formally determinate activity cannot be, constitutively and genetically, explained in material terms, and consequently neither can the capacity for acquiring the ability whose exercise counts as a formally determinate act or activity (e.g. squaring a natural number). If this is the case, then the practices themselves that give rise to and are constituted by formally determinate activities are inaccessible to and cannot be the outcome of beings who are endowed only with purely material capacities. It is in this sense that I believe our investigation blocks the project of naturalization, for it shows that the capacity for mathematics, logic, and any other formal activity cannot be constitutively and genetically explained in material terms alone, while also shedding light if only partly on why no other species on earth past or present has been able to generate and engage in the sort of formal practices that human beings have been able to.

7. THE ARGUMENT: KRA

I have said enough (perhaps too much) about the conceptual underpinnings of the present work. Before getting into the thick and thin of the argument, let me briefly state the basic structure of the overall argument. KRA can be represented syllogistically as follows:

(A) Nothing physical can determinately realize a formal operation.
(B) Some human activity determinately realizes formal operations.
(C) Some human activity cannot be (wholly) physical.

By "some human activity", I naturally mean the sort of human activity that issues forth from formal understanding: a distinct array of abilities (e.g. a logician's ability to prove Gödel's incompleteness theorems) whose exercise results in the relevant formal activity (e.g. actually proving the theorems). So neither the abilities nor the resultant activities can be (wholly) physical. The argument shows first that the physical cannot intrinsically realize any definite formal operation and that, therefore, whatever (if anything) intrinsically realizes a definite formal operation cannot be entirely physical. In chapters 4 and 5, I then proceed to argue that a certain kind human intellectual ability and activity intrinsically realizes definite formal operations, and therefore cannot be entirely physical (it cannot be entirely non-physical either). The reason why the physical cannot partake in formal understanding is because such understanding is determinate in ways that the physical cannot be.[68]

---

that is, no way of expressing and conceiving of quantities as discreet. So, a Pirahã mother does not know *how many* children she has, though she knows her children. Whether this is so is an empirical question, but that it is possible that there be a language that lacks numeracy seems to hold. In this sense, such a language could constitute a limitation to the acquisition of mathematical understanding (e.g. arithmetic, number theory, etc.), though there does not seem to be anything that would *in principle* impeded the introduction and emergence of numeracy in the language.

[68] The reverse is also true: the reason why formal understanding cannot partake wholly in physicality is because the physical is determinate in ways that no kind of understanding can be. Ross calls this 'the transcendent determinacy of the physical' as opposed to 'the formal determinacy of thought/understanding' (Ross, 1992, 2008). We will address this in chapters 4 and 5.

2
Kripke and The Formal Indeterminacy of the Physical

What follows shall consist in the examination of two main claims: first, the claim that Kripke's unpublished argument against computationalism (Kripke 1984a & 1984b; see Buechner 2011) *can* be formulated as an argument for the *formal* indeterminacy of the physical;[69] second, the claim that this argument is *sound*. I shall argue in defense of both. First, I will show that Kripke's unpublished argument against computationalism entails the formal indeterminacy of anything physical. Thus, unlike Kripke's, our incursion into computationalism and functionalism can be considered a means to an end rather than the end itself.[70] Computationalism (and functionalism), in the philosophically interesting ways we will canvass, will stand refuted *as a metaphysical theory of formal understanding and activity* if it is the case that the physical cannot realize *determinately* any formal (mathematical, logical) operation. The adverb 'determinately' is crucial to properly understand what that claim really amounts to and we will elucidate the concept of 'formal determination/determinacy' in this chapter. I have in mind *realism* about physical computation[71]—the view that physical processes can compute mathematical and logical functions *intrinsically*, i.e. in a way that is wholly independent of human intentions, conventions, interpretations, etc.—and not non-realism about physical computation—the view that the physical can only compute mathematical and logical functions in a manner that is wholly or partially *extrinsic*, i.e. dependent of human intentions, conventions, interpretations, etc. Now suppose you hope to explain, not merely model, the ability of human beings to compute mathematical and logical functions in terms of the realist conception of physical computation of mathematical and logical functions (that is, in terms of the way in which the physical can be said to compute mathematical and logical functions). Then you are the target of Krikpke's argument. For it shows that realism about physical computation is untenable, for physical computation cannot, in the full determination of its formal properties, rid itself of observer-dependence—a conclusion that would lead to absurdity if applicable to our ability to compute and apply mathematical and logical functions and rules (e.g. to add, to conjoin, etc.). The Kripkean argument supports a non-realist view

---

[69] The Kirchberg-am-Wechsel lecture – titled "Lessons on Functionalism and Automata" – was presented at the International Wittgenstein Symposium on August 23, 1984. There is also a tape recording in the Kripke Archives (CUNY) of another anti-functionalist lecture, given at Duke University on April 1, 1984, on occasion of the Patterson Lecture (See Buechner 2011, p. 344 n.3). The only detailed discussion of Kripke's anti-functionalist lectures is a paper publishd by Jeff Buechner (Buechner 2011). In what follows I will be relying on both Kripke's Kirchberg-am-Wechsel lecture and Buechner's detailed discussion of the lectures. It seems neither Kripke nor Buechner were fully aware of the argument's implication against the determinate physical realization of mathematical (and logical) functions. There is also an unpublished paper against functionalism by Kevin J. Lande (Lande 2011) in which the thesis that physical facts and laws do not suffice to determine which function a physical object computes or realizes is defended. Some parts of this chapter build on that paper by Lande. It is strangely similar to Kripke's argument. But Lande does not take (in my opinion) the argument against functionalism to its full logical consequence: that if computing an abstract function *f* is possible, then the computation of *f* cannot be an entirely physical phenomenon. But then again, neither did Kripke. Physical computation would have to be something other than the computation of abstract (e.g. mathematical or logical) functions (for an account of physical computation along these lines, see Piccinini, 2017, 2015). But then physical computation cannot be the kind of computation we are in engaged in when computing mathematical and logical functions. This claim will become fully clear by the end of the next chapter.

[70] Recall the caveat we gave at the outset of chapter 1. To paraphrase Kripke, the reader might consider treating following not so much as an exposition of Kripke's views but of 'Kripke's argument as it struck Antonio, as it presented a problem for him'.

[71] For this seemingly idiosyncratic but rather appropriate use of 'realism' in philosophical discussions about concrete, physical computation (not to be confused with the way 'realism' is used in the philosophy of mathematics, for instance, which could be tantamount to 'Platonism' or 'Meinongianism'), see (Ladyman, 2009).

of physical computation, but so does Putnam's (Putnam, 1988) and Searle's (Searle, 1992) both of which fall under the same category but for entirely different reasons. Kripke's argument is not a triviality or pancomputationalist argument, as we shall see. Kripke's argument hinges on issues of indeterminacy, specifically whether the physical in any given case can in virtue of any physical property (or set of suitably organized physical properties and components) alone suffice to determine a definite, unique mathematical or logical function so as to exclude incompossible mathematical or logical functions and contradictory results from obtaining at the same time and in the same respect. Kripke's argument goes through even if we assume, for the sake of argument, that triviality and pancomputationalist problems have been resolved.

As will become clear in what follows, we argue that if Kripke's argument is indeed sound, it may very well require us to re-examine our accounts of the nature of human intellectual (e.g. mathematical) activity if it is the case that *that* sort of activity genuinely realizes formal operations (Chapters 4 and 5). This chapter is organized into 4 sections. In the first section, we give a brief exposition of the infamous sceptical paradox developed by Kripke which the source for the reflections that later led to his anti-computationalist argument. The skeptical paradox, as we present it here, differs in scope from Kripke's original presentation in that we stick in its formulation to mathematical (and logical) rules and functions and do not venture into its implications for meaning. The second section will give a detailed exposition of the concept of computation relevant for the explanation of formal understanding and activity and will survey some of the main differences between computationalist, functionalist, and computational functionalist approaches to the cogitative. This marks the first half of the chapter. The third section presents and partially defends Kripke's unpublished argument against computationalism. The fourth and final section outlines some of the constraints that any concept of computation must respect if it is to intelligibly figure in an explanation of human formal activity, namely *our* computations of mathematical, logical, and other (say, grammatical) functions. This chapter will attempt to show that Kripke's anti-computationalist argument entails that no physical mechanism, state, and process can determinately realize a pure function, that is, compute a mathematical or logical function in any objective and human-independent way (whether this be logical operations and pattern of inferences like conjunction and *modus ponens*, elementary arithmetic operations like the addition and the successor functions, algorithms for computing a logical or mathematical function and generating well-formed grammatical, i.e. morpho-syntactic, expressions, etc.). The argument developed here paves the way for at least two of the most important philosophical implications of Kripke's argument against computationalism, namely (1) that there are no formal operations (that is, no *objective* application or following of formal rules for the computation of mathematical and logical functions) in physical mechanisms, and (2) that human intellectual activity if it genuinely realizes logical or mathematical operations cannot be (wholly) material; or contrapositively, if our intellectual activity is wholly physical, then we never genuinely realize any logical or mathematical operation (this will we show in chapter 4). Though the present examination of Kripke's argument goes well-beyond anything Kripke explicitly said in his unpublished lectures, such consequences can be readily shown to follow from his argument. Let's proceed.

1. THE ORIGINS: RULE-FOLLOWING AND THE SCEPTICAL PARADOX

Let us begin by considering the distant origins of Kripke's argument against computationalism. In (Kripke, 1982, 1981), Kripke develops what he calls a skeptical "paradox" which says that there is no fact of the matter as to what our linguistic symbols mean and, therefore, no fact of the matter as to

which rules are being followed when such symbols are used.[72] Both 'third-person' facts (say, pedagogical, neurological, behavioural, dispositional, etc.) and 'first-person' introspectable facts (e.g. mental images, sensations, past intendings, etc.) are deemed insufficient to determine a definite rule amongst incompossible ones for the use of an expression. All of the facts are compatible with incompossible rules or functions. To drive home the point, Kripke introduces a mathematical function called 'quaddition' (Kripke, 1981, p. 243, 1982, p. 9ff) whose symbol is '⊕' and which he defines as,

$$x \oplus y = \begin{cases} x+y, \text{ if } x, y < 57 \\ 5, \text{ otherwise} \end{cases}$$

For any pair of numbers less than 57, the quaddition function yields the same values as the addition function. For any pair of numbers greater than 57, the quaddition function yields 5, whereas the addition function yields their sum. For any numbers less than 57, though the outputs are the same, any one will constitute a sum in the case of the addition function, whereas any one will constitute a 'quum' in the case of the quaddition function. Now, formally, both functions are determinate and distinct. So this is not an issue about mathematics.[73] The problem concerns, rather, the *comprehension* and *application* of the relevant mathematical rules, e.g. the computation or implementation of the functions. What determines whether what has been applied in a given case is addition rather than quaddition or quaddition rather than addition? Suppose now that only physical (and even phenomenal) states, processes, and changes can count as the sort of facts relevant for the computation or application of mathematical functions or rules.[74] Then, if you really compute the addition function or the quaddition function, then there must be some facts or features about you (e.g. education background, "mental history", brain processes, behavior, dispositions, mental images and samples, etc.) that determine which function you've computed. Assuming, for the sake of argument, that numbers greater than 57 have never been calculated or even that numbers greater

---

[72] The literature on Kripke's skeptical paradox is too vast to be reasonably surveyed here. Some notable mentions are (Ahmed, 2007; Anscombe, 1985; Bloor, 1997; Boghossian, 1989; Hacker and Baker, 1984; Kusch, 2006; Mcdowell, 1992; McDowell, 1984; Miller, 2007 ch. 5 and ch. 6; see papers in Miller and Wright, 2002; Searle, 2002, pp. 251–264)

[73] This is why philosophical issues raised by formal undecidability (e.g. the Lucas-Penrose Gödelian arguments) and non-standard models of first-order logics (e.g. the Lowënheim-Skolem theorem) are not the same as the *philosophical* issues raised by Kripke's formal indeterminacy argument. We will deal, albeit briefly and only in sketch, in chapter 5 with an objection that uses, confusedly, Skolem's theorem against KRA. A full-blown paper can be written only on the philosophical significance of Gödel's and Skolem's theorems for KRA.

[74] I, not Kripke, is placing this constraint. It does, however, closely correspond to his use of third-person and first-person sorts of facts. At any rate, such a constraint should seem acceptable to a naturalist or physicalist. As we made clear in the introduction, the notion of physical we are using here is much broader than that of the physicist (though it includes it) and connotes not only the material properties of the world at the level at which physics and chemistry investigate them, but at the level of biology, neurology, geology, and so forth. We can even include, as Aristotelians would, sensation and (non-human) perceptual consciousness as being an entirely material phenomenon, perhaps emergent but irreducible to lower levels of physical explanation. Our argument will apply all the same, and so does the sceptical paradox (see e.g. Kripke, 1982, pp. 14–15 where Kripke points out that the crucial differences between Quine's indeterminacy of translation arguments and Wittgenstein's skeptical argument is that Wittgenstein did not limit the type of potential determinants of meaning to behavior but rather imposed no limits on what kind of fact could count as a potential determinant, allowing for an "extensive introspective investigation" to see whether any "internalist properties" could suffice). Note also that the concept of 'conscious' is, in analytic philosophy, intertwined with the notions of 'subjectivity' and 'phenomenology', the latter one being understood in terms of the qualitative aspects of experience (e.g. qualia). So the notions of 'phenomenal' and 'phenomenology' at play here are far removed from the Husserlian concept of phenomenology, though things have begun to change recently. For more, see (Gennaro, 2018; Garvey, 2011; Jackson and Smith, 2007; Zelazo et al., 2007; McLaughlin et al., 2009).

than 57 cannot be calculated—say, let '$n > 57$' be a placeholder for any number too great to be computed in a finite amount of time and in a finite amount of steps (and there surely are such numbers)[75]—it seems that any application of the rule of addition can count just as well as an application of the rule of quaddition, and vice versa. But then there is accord and conflict at the same time and in the same respect. Hence, nothing determines which rule is in fact being applied or followed.

Suppose, for example, that the function you've been calling 'addition' all along is actually the quaddition function and that this is true of you and your community's use of the word 'addition'. Then whenever you intend to 'add' you would really be intending to quadd.[76] But what determines that this is *not* the case (or that it *is* the case)? Recall that quaddition yields, for any numbers up to 57, the exact same values as addition, incompossible values otherwise.[77] Suppose you are a physicalist, whether of the type or the token sort. What could physically determine that a given case of calculation is a case of addition and not a case of quaddition and, therefore, that one intends one rather than the other? Appeal to symbols and causal relations between symbols will not do. For symbols are interpreted and conventionally determined, and the use of '+' in computing numbers less than 57 can be made out to accord with the quaddition function, just as the use of '⊕' in computing numbers less than 57 can be made out to accord with the addition function. Similar considerations purportedly apply to the *words* 'addition' and 'quaddition' whether written, or spoken, or heard, or pictorially represented in one's mind. This is why to respond to the skeptic by saying, 'How do I know that I add? Because I just do!" won't do even if true because your use of the word 'add' can be given a non-standard interpretation (e.g. unbeknownst to you, by 'add' you've really meant *quadd* all along) that is wholly compatible with everything you believe, learned, intend, and

---

[75] '57' is just a case in point, since there is, and always will be, a number which someone or no one has never computed (and infinitely many which no one ever will—say numbers too great to be computable in a finite amount of time). We can always make that number greater. But we can let, for the sake of simplicity, '57' stand for such number.

[76] Bear in mind that, as Kripke remarks, there is a difference between 'thinking (or being under the impression that) one is following a rule' and 'actually following a rule' (see, e.g., Kripke 1981, p. 283; cf. Wittgenstein *PI*, §202). The fact that you were under the impression that you were adding does not necessarily make it the case that you were adding. Otherwise, there would be no difference between 'taking oneself to be following a rule' and 'actually following a rule', and thus no intelligibility to the possibility of mistakes, of confusion, of illusion or delusion, and of a failure of understanding (or of application). If one is a physicalist, whether of the 'type' or 'token' strand, then one would have to point to a physical or material fact (or a set of such facts) that could determine (whether by identity, supervenience, realization, role satisfaction, or what have you) which function amongst incompossible ones is being carried out when a formal operation takes place—at least if she hopes to maintain that pure functions can be, and have been, physically computed. Notice that the problem is not merely epistemological. It is asking what sort of (e.g. physical) fact or feature could possibly determine the application of a definite, unique function.

[77] Note that due to the entrenchment of predicates, one can define addition in terms of quaddition and vice versa, analogous to the way green can be defined in terms of grue and vice versa (Goodman, 1983) . Hence, *inter alia*, objections of the sort 'addition is more fundamental', 'addition is more natural', 'addition is the simplest hypothesis', 'quaddition presupposes addition' miss the point. Consider: twin-Earthians, who have been doing quaddition rather than addition, could respond much the same way, 'What Earthians call 'addition' presupposes quaddition and is nothing more than quaddition up to numbers less than 57, something else after that'. So, why one function rather than another? Who is in the right and what determines who is in the right? Humean past regularity is not an option. For, as the Humean herself would concede, past regularities accord with incompossible functions and therefore do not suffice to determine potential and counterfactual cases in any objective way. Because past, habitual regularity conforms with incompossible functions (e.g. addition and quaddition), and all there is (according to the Humean) to computing correctly a given function is exemplifying some past, habitual regularity, it follows that it is indeterminate which function (if any) one has been or is following. Pragmatic considerations, of the sort Quine would recommend, are totally acceptable when it comes to deciding which function we will take someone or some physical system as computing. But they are irrelevant for the question about whether any non-observer-dependent phenomena, in particular any type or token of physical phenomena, (can) determine a unique, definite function.

have done with respect to what you call 'addition'. Plus, the skeptic is after a fact or feature that determines that you are actually adding, rather than quadding. So the question is, what kind of fact determines that you add rather than quadd? And the same question is raised with respect to intending to one function rather than some other incompossible one.

To respond by saying, 'well, if I could or I were able to compute numbers greater than 57 so as to continue to do the same thing I've been doing all along (say, when using the symbol '+'), I would *add* those numbers' is of no avail. For, this is tantamount to the circular and trivial statement 'well, if I were able to compute numbers greater than 57 and to continue to add them, I would add those numbers'. Yet what have you been doing all along is precisely what is in question. Mental images, like symbols, are equally mathematically indeterminate since they can be made out to accord with incompossible functions. Educational and performance history too fall prey to similar considerations as they consist of a finite sample of computations or applications of mathematical rules, and any finite number of applications of a given rule can be made out to accord with the finite application of some other incompossible rule. Same goes, interestingly enough, for intentions—e.g. "…ultimately, if the sceptic is right, the concepts of *meaning* and of *intending* one function rather than another will make no sense. For the sceptic holds that no fact about my past history - nothing that was ever in my mind, or in my external behavior - establishes that I meant plus rather than quus. (Nor, of course, does any fact establish that I meant quus!)" (Kripke, 1982, p. 13) The important thing to bear in mind is that the facts appealed to throughout the skeptic's argument as possible determinants of the formal aspects of the activity are the so-called third-person and first-person (introspectable) facts. The former set of facts generally encapsulates public, shared, and (in principle) observable physical or behavioral phenomena about an individual, whereas the latter encapsulates private, non-shared, and indirectly observable (or not observable at all) phenomena such as mental images, non-manifested intentions, feelings, imaginings, etc.

Let me make a point about 'intending' that is somewhat different from that of Kripke's and more relevant to the anti-computationalist argument to be developed. It is certainly the case that 'intending to follow a formal rule', that is, having the intention to compute some function '*f*', might not always be sufficient for determining whether *f* or any function at all is what is being intended (and, again, for the physicalist, we would need an account in terms of properties of matter sufficient for the determinacy of the formal content of *that* intending so as to block incompossible formal contents). For even if I intended to apply some formal rule *R*, that still does not mean that I actually applied *R*. I could've made a mistake or simply been confused about what the formal rule *R* is. A case of total misunderstanding, conceptual confusion, or delusion could lead to me to intend the application of what in reality is no function or rule at all. Therefore, what I do or end up actually doing need not be the same as what I intended to do, even though I *can* many times carry out what I intend to carry out. So, the question is whether there is any explanatory role that intentions (can) play in determining the formal rule being applied or, if none is actually applied, whether intentions can be determinate with respect to mathematical and logical rules and functions. We know that this should be answered in the affirmative, the question is on account of what? The answer to this will differ depending on which metaphysical framework underpins one's explanatory standards concerning possible explanations of the nature of intentions, and in particular intentions with respect to the computation or application of determinate mathematical/logical functions or rules. If physicalism is true, then any 'intending' will be physical and given the sceptical paradox it would seem impossible to intend the application of a determinate formal rule.[78]

[78] Though Kripke does not quite take this route, I would say that intending can be *formally* determinate only if it is the intending of a formal rule *which has been understood*. The determinacy of one's intending to compute a function '*f*' is consequent upon one's having understood (what it is to compute) *f*, to whatever degree. But then the question arises, on

Notice that the skeptic is after a fact *about the individual* that can determine which rule she followed in the past (if any) and is following presently (if any). All your *past* computations comply with either function[79] and your present computation can be likewise made out to 'accord' or to 'conflict' with incompossible rules—whether addition or quaddition or still some other incompossible function (e.g. skaddition $=_{df.}$ $x + y$, if $x,y < 57$; $2xy(x + y)$ if $58 \geq x,y \leq 10^{450,328}$; 5 otherwise). Moreover, if we let '57' mark the numerical threshold above which numbers are too great to be computed by anyone in a finite amount of time and in a finite number of steps, all *past, present* and *future* computations (which will be finitely many) will comply with indefinitely many incompossible functions.[80] Hence, concludes the skeptic, there is no fact of the matter as to whether anyone adds or quadds.[81] Notice also that this argument is a *constitutive* or *ontological* one, not merely a *cognitive* or *epistemic* one. The point is not that there are some facts that escape us and therefore that we cannot determine, epistemically, which function is being applied or computed. This epistemological version of the argument would be equally devastating for a computational or a functionalist account of our higher cogitative powers (more on chapter 5). The point is, rather, that

---

account of what is it determinate that a definite function can or has been be understood and intended? Our task will be to show that there is no intelligibility to a physicalist answer to this question.

[79] Kripke's original formulation (Kripke, 1981, 1982) of the skeptical paradox consists, partly, of what I will call a *tensed* account of the determinacy of rule following or of rule determination. "Although I myself have computed only finitely many sums in the past, the rule determines my answer for indefinitely many new sums that I have never previously considered. This is the whole point of the notion that in learning to add I grasp a rule: my past intentions regarding addition determine a unique answer for indefinitely many new cases in the future." (Kripke, 1982, pp. 7–8) Notice the (fallacious) shift from '*the rule* I learned in the past determines the correct answer for indefinitely many, and as of yet computed, arguments' to '*my past intentions* regarding the rule determine the correct answer for indefinitely many, and as of yet computed, arguments'. Kripke's tensed sceptical paradox would be best expressed by means of the question, 'can I do now, presently, precisely the very same thing (i.e. the very same computation) that I did in the past finitely many times?' That is to say, is there a fact *about oneself* that determines whether one's present calculation or computation *is the same* as that which one has learned and carried out in the past finitely many times? The question, 'which rule did she learn and internalize?' has no determinate answer, says the sceptic, if all the characteristics of the facts appealed to are historical, physical, behavioral, dispositional, introspectable, and so forth. A non-standard interpretation of what is claimed to have been learned or done can always be given (for this very point, see Kripke, 1982, pp. 15–17).

[80] I side with (Hacker & Gordon, 1983) in thinking that it is extremely odd and strange to take one's carrying out of a function, one's following of a rule, as a matter of one's acting in accord *with past performances and intentions*. Past performances and intentions, let alone past states, are just irrelevant for following a rule (as is exemplified, for instance, when one learns and understands a rule *for the first time* and proceeds to carry out a formal operation in accord with it, or when one simply proceeds to follow an arithmetical rule like division; we don't consult what we did yesterday). But, as it turns out, the focus on *past* happenings is not essential to the argument. For the skeptic's argument can be made, and *is* in fact made (Kripke, 1982, p. 21ff), with respect to any occurrent and potential application of a formal rule.

[81] See (Boghossian, 1989; cf. Searle, 1987, 2002, pp. 251–264; but see Kusch, 2007 for a critique of Searle's argument). Both Hoffman and Searle argue that the skeptical paradox is absurd, albeit for slightly different reasons. Searle believes that any argument that entails that we do not mean anything determinate by our words and expressions must be taken as a *reductio ad absurdum* of the premises that lead to such a conclusion (cf. Kripke, 1982, p. 57). Hoffman (Hoffman, 1985) argues that, if Kripke's skeptic were correct, we would not be able to formulate and even understand the skeptical paradox itself; it is a performative contradiction and logically incoherent. I think both critiques are on the right, but we offer a different solution or diagnosis though we follow Hoffman's path. With respect to physical (and phenomenal, introspectable) facts and properties, the skeptical paradox is perfectly coherent. Indeed, if all there was to human formal understanding and activity was just physical processes and states, then the skeptic's argument would indeed have some bite: there simply could not be any such thing as semantic and formal determinacy. But, of course, since we do apply determinately definite formal rules, and we do (and can) acquire formal understanding, it follows that formal understanding and the activity of applying or following formal rules cannot be (wholly) a matter of physical (and phenomenal) processes. It also follows that intending to do so (say, intending to add or carry out a formal operation) in consequent on understanding the formal rule or function which is being intended and therefore the intending cannot be wholly a matter of matter because it must be, at least content-wise, determinate with respect to the relevant formal concepts and rules. See chapters 4 and 5.

all of the types of facts (i.e. third-person facts and first-person introspectable facts) about any arbitrarily selected individual person that can be appealed to as possible determinants of the definite, unique function being computed by a given individual will be equally compatible with the computation of an indefinite number of incompossible functions, to the effect that nothing about the individual herself can suffice to determine a unique, definite mathematical or logical operation. Indeed, since nothing determinate is fixed, there is no such thing as the objective, determinate computation of a mathematical function (Kripke, 1982, p. 21). So there is no such thing as the determinate *grasp* and *application* of mathematical rules in any intrinsic and objective sense. There is nothing about us in virtue of which it is determinate that we compute a definite mathematical function and, by extension, nothing about us in virtue of which we can reason determinately in logical forms of inference. All purported determinant facts are compatible with incompossible rules. Not even an omniscient God, or an epistemically idealized observer, would be able to tell which function is being applied given third-person (e.g. neural connections) and first-person facts (e.g. memories of past intendings).[82]

In the remainder of this chapter, we will argue that this claim is true but only in a qualified sense. That is to say, it is true that there cannot be such a thing as the determinate application of mathematical (and logical) rules if the only category of possible determinants for the application of a definite formal rule is taken to be (a subset of) the category of physical (and phenomenal) structures—whether cashed out in terms of physical states, events, properties, processes, or mechanisms.[83] The fallacy in the skeptic (if it can be called a fallacy) is what I shall call 'the naturalistic fallacy', namely the presumption that the only ontological category that could plausibly provide the determinants of the formal features of the cogitative is the category of material structures, regardless of whether their power to determine formal features is cashed out reductively or non-reductively.[84] The sceptical paradox brings home the point that the sorts of facts that are generally considered in mainstream analytic philosophy of mind, i.e. third-person and first-person facts, are insufficient for the formal structure of the cogitative.[85] Given the category of facts that

---

[82] "So whatever 'looking into my mind' may be, the sceptic asserts that even if God were to do it, he still could not determine that I meant addition by 'plus'." (Kripke, 1982, p. 14) "When the sceptic denies that even God, who knows all the facts, could know whether I meant plus or quus, he is simply giving colorful expression to his denial that there is any fact of the matter as to which I meant." (Kripke, 1982, pp. 40–41) (see also Miller, 2007, pp. 165–201) Replace "which I meant" with "which I followed/applied" and you get the formal version of the argument. If meaning is to be construed in terms of rule following, then evidently our version of the argument will bear on any such theory of meaning. But even if meaning is to be explained in terms of use and therefore in terms of the application of the (semantic) rules of the language, so to speak, it remains doubtful that *use* would be univocal with the application of a (set of) formal rule(s) for the computation of mathematical or logical functions. If it is so, however, then our argument will apply just as much to such an account of language and linguistic understanding (see chapter 6 for the pursuit of this line of reasoning).

[83] We will not develop the argument in detail with respect to the phenomenal and the subjective, as this is generally understood in analytic philosophy. Kripke (1981, 1982) considers all such facts and finds them wanting. Our discussion of his anti-functionalist argument (Kripke 1984, Buechner 2011) will illuminate Kripke's previous considerations. It should be relatively easy, however, to see how our argument applies to such things as qualia, mental imagery, sensory impressions, behavioral dispositions, and so forth.

[84] Note that first-person facts such as the phenomenal aspects of consciousness are equally unable to determine a definite mathematical or logical function. So, panpsychists and dualists about phenomenal consciousness face the problem that neither the physical nor the phenomenal can *suffice* to determine the definite formal features and aspects of the cogitative since any phenomenal fact is compatible with incompossible formal rules or functions.

[85] By third-person facts we mean, of course, facts that are in principle observable yet intrinsic to the person, such as their physical behaviour, their dispositions, their neurobiological mechanisms, etc. The "sceptical solution" (Kripke, 1982, pp. 55–113) to the skeptical paradox could be said to appeal to 'third-person' facts (i.e. assertability conditions) in the sense that they are 'observable', public, and shared; but in the case of the sceptical solution, such facts are not intrinsic to any individual person. They are (intrinsic?) to the community which determines the conditions under which it can be asserted that some pattern of behaviour manifested by an individual satisfies the relevant conditions (say, for carrying

naturalism allows, the use of the symbol '+' and the word 'addition' finitely many times cannot exclude them being symbols that refer to the quaddition function instead, just as a Twin-Earthian's use of the symbols '⊕' and the word 'quaddition' finitely many times would not be able to exclude the addition function as the referent of such symbols (granted, for the sake of argument, that no operation has been carried out over any pair of numbers greater than 57 or that '57' stands for a number too great to be computed in finitely many steps in a finite amount of time). All third-person facts and first-person introspectable facts seem to accord with either function, yet the functions are mathematically or logically incompossible and thus impossible to satisfy at the same time and in the same respect. Hence, there is no fact of the matter as to which (and as to whether any) function is being carried out and thus no fact of the matter as to what the relevant symbols or words mean.

As far as the purposes of this chapter are concerned, this rough sketch of the sceptical paradox is more than sufficient. We will no longer be concerned with the meaning of expressions except insofar as it concerns fixing an interpretation that would allow a physical item to become a *formal* symbol and express or represent a given formal operation (e.g. '$x + y$'). Other than that, our focus will be on the ability to carry out formal (e.g. mathematical) operations, on the very act of following or applying of formal rules, on the *structure* of the act itself. Before moving on, however, it will do us well to pause and consider an aspect of the skeptical argument that Kripke considers fundamental and which, albeit in a somewhat different way, will come up in our Kripkean argument against naturalizing formal understanding. There is a *normative* or *justificatory* requirement imposed by the skeptic on any response to the sceptical paradox. So Kripke tells us,

> The basic point is this. Ordinarily, I suppose that, in computing '68+57' as I do, *I do not simply make an unjustified leap in the dark*. I follow directions I previously gave myself that uniquely determine that in this new instance I should say '125'. What are these directions? By hypothesis, I never explicitly told myself that I should say '125' in this very instance. Nor can I say that I should simply 'do the same thing I always did,' if this means 'compute according to the rule exhibited by my previous examples.' That rule could just as well have been the rule for quaddition (the quus function) as for addition … An answer to the sceptic must satisfy two conditions. First, it must give an account ofwhat fact it is (about my mental state) that constitutes my meaning plus, not quus. But further, there is a condition that any putative candidate for such a fact must satisfy. It must, in some sense, show how I am justified in giving the answer '125' to '68+57'. The 'directions' mentioned in the previous paragraph, that determine what I should do in each instance, must somehow be 'contained' in any candidate for the fact as to what I meant. Otherwise, the sceptic has not been answered when he holds that my present response is arbitrary. [my emphasis]
>
> (Kripke, 1982, pp. 10–11)

So consider what happens when one tries to carry out an operation over a pair of numbers greater than '57' (on the supposition that one has never computed numbers greater than 57). Despite which answer I produce, whether the answer is *correct* will depend on the rule or function being applied or computed. But the whole point is that not fact determines which function is being computed. So if

---

out addition); yet, intrinsically speaking, there's nothing about the individual that determines this. In isolation, no individual can follow a determinate rule. Different communities with different assertability conditions can equally well take the same pattern of behaviour manifested by given the individual to have satisfied the conditions for carrying out their mutually incompossible functions (say, addition and quaddition). There's nothing more to rule following than this, says the solution.

it's indeterminate which formal rule I've been and am applying, how am I *justified in virtue of the rule* in giving the answer that I give and how can it be determinate whether the answer I give is correct or not? Or is it all nothing more than a leap in the dark, an arbitrary choice, a blind application?[86] One response to the sceptical paradox which Kripke considers at length is *dispositions*. What one is disposed to answer when calculating the pair ⟨57, 68⟩ will determine whether one is following addition or following quaddition. So our mathematical dispositions determine which rules we have been taught and compute. Kripke finds the dispositional response wanting. First, because it conflates descriptive issues with normative issues. Dispositions tell us how we will or would answer to a given mathematical problem, but it cannot tell us what we *should* or *ought* to answer to a mathematical problem. The standard of correctness that is the mathematical rule is not dispositionally determined, nor dipositionally tracked. For, secondly, one can be disposed to give the *wrong* answer and therefore if dispositionalism were correct, any answer we were disposed to provide would be correct which is absurd. Thirdly, dispositionalists have to assume what is in question: if it is indeterminate which function is being applied, then it is indeterminate whether we are disposed to answer *correctly*. For, one cannot talk about an answer being 'correct' or 'incorrect' without knowing the norm that the answer adheres to or deviates from, so without knowing which function is being computed. So it is a *petitio*. But, fourthly, dispositions, like other physical, mental, and behavioral facts, are compatible with incompossible functions since, at least with respect to infinite mathematical functions like addition and quaddition, our dispositions are finite and thus will agree with incompossible infinite functions like addition and quaddition. One can always make the differentiating threshold between the functions higher than '57'.[87]

The requirement that whatever we point to as a possible determinant of the function being realized must constitute a *normative* ground for the answers one gives to the relevant arithmetical problem is controversial. At any rate, there is a sense in which the rule constitutes a *justificatory* and *normative* ground for the answers one gives and if nothing determines which rule is being followed, nothing determines whether one is justified and correct mathematically in giving one answer rather than another. Thus, the determinant of a definite formal operation must make intelligible our talk of 'what I ought to answer if presented with the function '68 + 57'', not merely 'what I would be

---

[86] Two things. First, a mathematical function or rule like addition is an arithmetical *norm* and as such a standard of correctness for its application. Mathematically, whether an answer to an addition problem I give is correct or not is determined by the addition function or rule. This is one sense of normativity and legitimacy. But there is another sense, which seems to *compel* one to answer a certain way. If I wish to reason validly, for example, then I should reason such-and-such. But I can wish not to, of course, in which case there is not compelling force or 'should' (assuming it is not, in whatever sense, our duty as rational animals to aim to think validly; yet even here the 'ought' will be of the second rather than the first sense of 'normative'.) For example, if what I have learned and computed all along has been the addition function, then if I wish to accord with it, I *should* for any pair of numbers greater than '57' (e.g. ⟨57, 68⟩) answer with their sum (e.g. '125'), so I should add them. If what I have learned and computed all along has been the quaddition function (say, under the guise of the '+' symbol and the word 'addition'), then if I wish to accord with it, I *should* for any pair of numbers greater than '57' (e.g. ⟨57, 68⟩) answer with their quum (e.g. '5'), so I should quadd them. The addition function does not compel me to compute it. But if I wish to compute certain numbers in accord with the addition function, the I have to (should?) add. See n. 159. Just bear in mind that it is the first sense of normativity that is the one operative in the arguments to be developed in this monograph.

[87] Though Kripke does not consider the following option, the numerical threshold which manifests the difference between the two functions could likewise be placed on an altogether different *kind* of numbers, say complex or imaginary numbers, such that the quus-like function yields the exact same values as addition for any pair of real numbers, but altogether different values for any pair of complex or imaginary numbers. And since the real numbers are infinite, not even an actual infinity of computations over the real numbers will be sufficient to determine a unique, definite function. The result will be *mutatis mutandis* the same. We will show this when come to Stabler's objection to Kripke's argument.

disposed to answer if…' since I can be disposed to give the 'wrong'[88] answer (i.e. dispositions are fallible) or simply not be disposed to answer anything at all (e.g. the 'range' of our dispositions is finite whereas the domain and range of functions like addition and quaddition are infinite; hence, dispositions cannot cover all possible cases of an infinite function). As the preceding paragraph made clear, Kripke argues that dispositions won't do the job because they are *descriptive* rather than *normative* and the dispositionalist response confuses issues of *source* with issues of *legitimacy* (cf. Goodman, 1983, pp. 60–61).[89] Hence, the dispositionalist might be able to account for the causal, behavioral, and psychological source of the produced answers to a given mathematical problem but will be unable to account for the formal legitimacy, justification, and correctness of such answers. Dispositions do not suffice to determine the formal rule that is being or will be applied. This is what can be called *the normativity constraint* that must be met and which (so the skeptic claims) dispositionalism cannot meet.[90]

---

[88] I can also be disposed to give the 'correct' answer without any competence. I'm just so disposed. Of course, we shouldn't even talk of 'wrong' or 'correct' answers at this moment since whether a given answer is wrong or correct is tantamount to whether the answer accords or conflicts with the rule being carried out. The correctness or incorrectness of a particular case depends on the rule or function in question (cf. Kripke, 1982, pp. 28–30).

[89] My concerns are a bit different from Goodman's. Goodman holds that deductive (and inductive) inference rules are justified by their conformity to accepted deductive (and inductive) practices, that is, by the particular deductive (and inductive) inferences "we actually make and sanction" (Goodman, 1983, p. 63). Particular deductive (and inductive) inferences are in turn justified by the deductive (and inductive) rules we deem acceptable. The circularity is virtuous, says Goodman, the point being that both rules and particular inferences are justified by being brought into agreement with each other (Goodman, 1983, p. 64). Whatever the merits of this position, Godman did take it to be the case that we could follow definite deductive and inductive rules; his point was rather that our application of a definite rule would be justified by its conformity to a rule we deem acceptable for whatever reasons and not by something other (e.g. unattainable guarantees that future predications, in the case of induction, would be correct) than acceptable inferential practices. The skeptical paradox concerns the intelligibility of following any formal rule, and our Kripkean argument against functionalism shows that physical phenomena cannot suffice to determine a definite formal rule and therefore that wholly physical mechanisms cannot on account of their physical properties alone follow any formal rule. It follows that if we are wholly physical mechanisms, then we cannot even act in conformity with accepted rules of inference since there would be no such thing as conforming to a definite rule.

[90] For a useful and thorough treatment of the 'normativists vs. anti-normativists' debate, see (Kusch, 2006) Causal, dispositional, and counterfactual answers to the skeptical problem are closely related and all fall prey to the same objections. We will see this more clearly in the context of Kripke's anti-functionalist argument. But let me adapt Kripke's considerations about normativity to make a point that is fundamental to the Kripkean argument I am putting forth. Consider: what I ought to answer in a given case depends on what I am doing, i.e. on what rule I am following or applying (e.g. if quaddition, I ought to answer '5', even if I'm disposed to answer '125'). So, an account of how one ought to proceed in a particular case given a mathematical or logical problem will depend on which rule is being followed and applied (and, of course, on whether I have the commitment to do it right or to proceed correctly). But that is precisely what is in question! To try to answer the skeptic by pointing to what I would answer or would've answered is both to miss the point and to conflate a normative issue with a descriptive, causal one. On the other hand, to try to refute the skeptic by simply pointing to what I ought to answer, or ought to have answered, or ought to have continued answering, is to assume precisely what is in question, for one can only determine what one ought to have answered if one has already determined what one is (or was) doing, namely which rule is being followed. Third, the fact or feature of the realizer which satisfied the normativity constraint must be dispositive for any possible case of the rule, viz., it must determine what the correct answer is not only for the actual case (e.g. the problem I am currently calculating, '57 + 68') but for any potential and counterfactual cases of the application of the very same rule or operation. Otherwise, if what one is doing is normatively indeterminate with respect to either potential or counterfactual cases, the actual 'doing' or application can always be made out to accord (or conflict) with an indefinite number of incompossible formal rules. This is *not* a slide into psychologism. We are not saying that the values of a function are determined by what we are doing such that the value that corresponds to '17 + 24' is '41' because of some psychological fact. The values of the function are determined independently of such facts. The function is a standard of correctness, a norm for its correct application and thus cannot be determined by its application. What we are saying is that what counts as a correct answer in a given case depends on which function that case is an instance of. So if I give an answer to an arithmetic or logical problem,

This is admittedly a rough summary of Kripke's anti-dispositionalism about rule-following (for the thrust of Kripke's arguments against dispositional analyses of rule following, see Kripke, 1982, pp. 22–37). Not all of his objections are equally strong. But this is not our concern at the moment. The dispositional response and Kripke's purported refutation of it has generated an immense amount of literature and controversy, and it is not our aim to weigh in on it.[91] Our concern here is to outline Kripke's skeptical argument about rule-following since therein lie the origins and precedents of Kripke's later argument against functionalism and computationalism. But Kripke's anti-functionalist argument has a life of its own and it is thoroughly anti-skeptical about our ability to follow formal rules, and so it cannot be entirely identified with the sceptical paradox. What is of relevance to us from the sceptical considerations is the following: we know that we (can) add rather than quadd and that we (can) quadd rather than add, and we grasp the distinction between these two formal operations. The question is what determines and accounts for the fact, metaphysically speaking, that we (can) do one rather than the other and that we (can) grasp one rather than the other.[92] The skeptic says 'nothing can' and so there really is no fact of the matter as to what rule is being applied and what rule is being grasped. But surely this is self-refuting (we show this in chapters 4 and 5). [93]

---

whether that answer is correct will depend on what I am doing, by which I mean 'which rule is being followed, which function is being computed'. In addition, we can talk about 'what ought to be answered' in that given case because we are not ruling out the (a) possibility of mistakes, of wearing out, of finite capacities, and of ceasing to exist, and because we are taking it for granted (b) that there is an occurrent intention to act in accord with a given rule. If we rule out (a), and maintain (b), then (ceteris paribus) I would be disposed both to answer correctly and to answer what I ought to answer, and thus my dispositions, i.e. what I would answer, now become an epistemically reliable (indeed, infallible!) guide to determining what I ought to answer in every single case, and thus what the correct answer is (since I would be systematically disposed to give the correct answer). But (a) is an idealization that is not, strictly speaking, true of anything physical. And (b) presupposes accord with a determinate rule, which is precisely what is in question. The idealization and abstraction have features that cannot be intelligibly ascribed to anything physical (if the thing we're ascribing it to is to remain a physical thing). On a slightly different note, I here diverge from Wittgenstein and Kripke. Though I agree with Wittgenstein that the *meaning* of word cannot contain, in some queer way, all the possible cases of its application, the same cannot be said about pure formal rules and functions. I have opted to talk about '*any* possible case' rather than '*all* possible cases' or '*every* possible case' to avoid giving the impression of an extensional commitment to actually existent *possibilia*.

[91] Some popular responses to Kripke's anti-dispositionalist arguments can be found in (Blackburn, 1984; Boghossian, 1989; Fodor, 1994, pp. 94–95; Horwich, 1995; Soames, 1998b) See also (Miller and Wright, 2002). For a fine bibliography on the vast literature on Kripke's rule-following considerations, see (Kusch, 2010).

[92] In line with (Kripke, 1984), I am taking the fundamental tenets of the skeptical paradox in a wholly different direction from that of (Kripke, 1981). Ours does not lead to any such paradox. The skeptical paradox, and thus the problem of formal determinacy for humans, obtains only if the only possible determinants of formal operations are physical (and phenomenal) facts. See chapter 4.

[93] The fact that we are talking about persons, about what we do or can do, when formulating the skeptical paradox is somewhat misleading. For, in our formulation, we are considering human persons only as purely material beings (with phenomenal properties like mental images)—a defective notion of the human person at best. If the reader feels a sort of uneasiness with our formulation of the skeptical argument, it might be because of this. For we are allowing only material or physical (and phenomenal) facts into the equation, and we are inquiring which (if any) can solve the problem of determining amongst incompossible pure functions which one is being realized (if any). A further point of clarification must be made. If you hold that the phenomenal is reducible to the physical, then our argument works. If you hold that the phenomenal, and the psychological in general, is irreducible to the physical, our argument will work equally well. For phenomenal properties like sensations, color perceptions, and so forth, are all equally formally indeterminate. (See, e.g., Kripke's (Kipke 1982) refutation that mental images or qualitative, phenomenal experiences can determine a unique and definite function. See also (Feser, 2013) for an argument to the effect that *meanings* or conceptual content cannot be determinately realized by phenomenal forms, say imaginings or mental pictures. The exact argument can be given from a formal, rather than semantic, standpoint).

Our strategy will be to prove a qualified version of the skeptic's answer, namely 'nothing *physical* can determine a definite formal rule or operation'. This much can be shown to follow from Kripke's anti-computationalist argument, and inasmuch as the skeptic's arguments against rule-following target physical and phenomenal properties as the possible determinants of the formal component of mathematical and logical operations, it can be shown to follow from the sceptical paradox too.[94] If we add the further premise that human cogitative powers are wholly physical, then we would be justified in saying that nothing determines metaphysically that a formal rule is followed and can be followed at all. We will show that the negation of the consequent of this conditional statement is true (chapters 4 and 5). For now, simply consider this: if the sceptic's conclusion seems fishy to you, that is because we *understand* these functions. We know what it is to add and what it is to quadd, otherwise we could not know what this talk of 'being indeterminate amongst such functions' amounts too. As I've already said, later on we will see that the skeptic's indeterminacy considerations cannot be coherently applied to a being who can understand such considerations—and who therefore can grasp, formulate, analyse, critique, bite the bullet with respect to, etc., such indeterminacy arguments, since this requires having a determinate grasp of the functions in question. Until chapters 4 and 5, however, we will limit ourselves to showing that no wholly physical mechanism can be, on account of material properties alone, a suitable subject of formal understanding since it cannot rule out, intrinsically speaking, incompossible formal rules or functions from being equally realized. Therefore, physical processes can never be a case of a determine, unique, formal rule or function. To see how this works, let us begin by considering the computational and functionalist theories of mind and how they purport to adequately explain formal understanding and activity, i.e. the cogitative.

2. WHAT IS COMPUTATION? RULE-FOLLOWING CONSIDERATIONS, FUNCTIONALISM, COMPUTATIONALISM, AND COMPUTATIONAL FUNCTIONALISM

Here is set of questions we want to address at this point,

Q1. What is a computation in the relevant sense of computing a mathematical or logical function? More specifically, what is it that we do when we implement or carry out a mathematical or logical computation)?

Q2. In which manner does functionalism, computationalism, or computational functionalism explain the computation of mathematical and logical functions? ('Functional computationalism' not being the label of a theory at all.)

---

[94] Such concepts as 'behavior', 'dispositions', 'intending or intentions', 'action', 'cause', 'mental images and representations', etc., as couched and understood within a physicalist framework (again, be it reductive or non-reductive) cannot be formally determinate. This will become obvious as we elucidate our argument concerning the formal indeterminacy of the physical. This means that if those concepts are to play a role in an explanation of human formal understanding, i.e. in the ability to engage in formal activity, then each and every one of these concepts must be excised of any physicalist innuendos if they are to intelligibly enter into an explanation of that sort of activity. It is not our argument that human behaviour and dispositions cannot be formally determinate, for example; rather, it is a consequence of our argument that these cannot be formally determinate if they are taken to be no more than purely physical processes, states, capacities, or what have you. So if human intelligent behaviour can partake in formal determinacy (as it surely can, otherwise teaching mathematics would be impossible), then it cannot be entirely physical nor explicable in entirely physical terms. We hint at this at the end of chapter 6, where we encourage a shift in conceptual frameworks for the explanation of behaviour, action, causation, etc., as they apply to human intelligent action and activity. A full treatment of this must, however, await another investigation.

The first question, Q1, is ambiguous. Are we talking about computation in the abstract sense in which it is studied in computability theory (a branch of mathematics) or as we talking about concrete computation? Note that it makes sense to say 'the computation took ten minutes' but also 'the computation is logically irreversible' or 'the computation is a case of logical transformation' (see Ladyman, 2009, pp. 376–377). The question that emerges is what is meant by 'concrete' computation? Take Alan Turing's successful attempt to hit upon a proper way to make mathematically precise the informal notion of an *effective procedure* for calculating the values of certain kinds of mathematical functions. Both the application of formal effective procedures and the complex task of making the notion of an effective procedure mathematically precise are instances of formal activity consequent upon formal understanding though not necessarily consequent upon the same formal understanding (Turing could both mathematize the intuitive notion of an effective procedure *and* carry out an effective procedure, say for adding, but most people can only do the latter which the nonetheless requires understanding what it is to fallow or apply the procedure in question). These are cases of concrete mathematical activity of a certain sort (i.e. having to do with computable functions), and all mathematical activity is concrete insofar as it is individual human beings who (have the ability to) *do* mathematics (e.g. linear algebra, computability theory, etc.). Thus, when Alan Turing carries out a computation, when he is computing a mathematical function, this is a case of concrete computation. But in the philosophical literature, concrete computation is taken to mean physical computation, that is, the *physical* realization of formally effective procedures or rules for solving a given mathematical or logical function. By effective procedure we mean an *algorithm*, that is, a set of finite formal rules that provide a way for calculating the values of a function in a finite number of steps. A function whose values can be calculated algorithmically is said to be a *computable function*. Thus, a function and an algorithm are not the same. For there can be more than one algorithm or effective method for calculating arbitrary values of a single function. The same is true of programming languages, a variety of which can express the very same algorithms and compute the very same functions. The naturalistic framework underpinning computational and functionalist theories of the mind leads them to insist that human concrete computations (and, more generally, human formal activity) is but the physical realization of abstract mathematical rules and functions (what we more or less refer to as 'pure formal structures' in chapter 1). Our adapted exposition of the sceptical paradox, particularly with respect to physical determinants of mathematical rules, should give the reader an intuitive picture of the argument to be mounted against the computationalist explanation of the cogitative.

Now, what is a computation? Like many other terms, 'computation' is spoken of in many ways. Let us avail ourselves of a rudimentary but immensely useful example of algorithm by philosopher Tim Crane:

> Consider, for example, a very simple algorithm for multiplying two whole numbers, *x* and *y*, which works by adding *y* to itself. It will help if you imagine the procedure being performed on three pieces of paper, one for the first number (call this piece of paper X), one for the second number (call this piece of paper Y) and one for the answer (call this piece of paper the ANSWER).
>
> Step (i): Write '0' on the ANSWER, and go to step (ii).
> Step (ii): Does the number written on X = 0?
> If YES, then go to step (v)
> If NO, then go to step (iii)
> Step (iii): Subtract 1 from the number written on X , write the result on X , and go to step (iv)

> Step (iv): Add the number written on Y to the ANSWER, and go to step (ii)
> Step (v): STOP
>
> Let's apply this to a particular calculation, say 4 times 5. […] Begin by writing the numbers to be multiplied, 4 and 5, on the X and Y pieces of paper respectively. Apply step (i) and write 0 on the ANSWER. Then apply step (ii) and ask whether the number written on X is 0. It isn't – it's 4. So move to step (iii), and subtract 1 from the number written on X. This leaves you with 3, so you should write this down on X, and move to step (iv). Add the number written on Y (i.e. 5) to the ANSWER, which makes the ANSWER read 5. Move to step (ii), and ask again whether the number on X is 0. It isn't – it's 3. So move to step (iii), subtract 1 from the number written on X, write down 2 on X and move to step (iv). Add the number written on Y to the ANSWER, which makes the ANSWER read 10. Ask again whether the number written on X is 0. It isn't – it's 2. So move to step (iii), subtract 1 from the number written on X, write down 1 on X and move to step (iv). Add the number written on Y to the ANSWER, which makes the ANSWER read 15. Ask again whether the number written on X is 0; it isn't, it's 1. So move to step (iii), subtract 1 from the number written on X, write down 0 on X and move to step (iv). Add the number written on Y to the ANSWER, which makes the ANSWER read 20. Move to step (ii) and ask whether the number written on X is 0. This time it is, so move to step (v), and stop the procedure. The number written on the ANSWER is 20, which is the result of multiplying 4 by 5.
>
> This is a pretty laborious way of multiplying 4 by 5. But the point of the illustration is not that this is a good procedure for us to use. The point is rather that it is an entirely *effective procedure*: at each stage, it is completely clear what to do next, and the procedure terminates in a finite number of steps. The number of steps could be very large; but for any pair of finite numbers, this will still be a finite number of steps.
>
> (Crane, 2016, pp. 61–62)

Two points of clarification seem appropriate at this juncture. First, when we talk about 'human computation' we mean the ability (underpinned by the capacity for that ability) of human beings to follow or apply formal rules determinately on account of their own abilities. The formal rules we will focus on for dialectical purposes are those whose application constitute effective procedures for calculating the values of a mathematical function, a logical function (e.g., the logical function 'XOR'), or some other effective formal operation. As noted in chapter 1, this methodological restriction to computable functions is neither essential nor necessary to our argument. But since the functions we shall be discussing are for the most part computable, and since the notion of computation as physically implemented has been generally taken to have paved the way for the naturalization of 'human computation', we wish to show that not even physical computation can account for human computation. But our argument is, of course, generalizable to formal understanding as such, which goes well beyond computability theory and computational implementation (e.g., computer science, software engineering, etc.). Secondly, given that our formulation of the skeptical paradox has been sort of a *précis* to Kripke's anti-functionalist/computationalist argument, the reader should have an intuitive idea of how the argument against computational theories of the mind, insofar as they avail themselves of a realist view of physical computation for the explanation of the cogitative, goes. Crane's algorithm is a useful and simple example that should make it readily intuitive how the process and what appears in the three pieces of paper if entirely physical can be *qu-d*, that is, made out to accord with incompossible (qu-like) algorithms. An algorithm for a mathematical function is on the same ontological footing as the mathematical function itself: algorithms are pure formal rules, falling thus

under the category of pure formal structures. In general, when it is said that physical systems compute mathematical or logical functions, it is actually meant that physical systems follow (i.e. realize or implement) some algorithm(s) for calculating the value of a certain function. Literally, this cannot be true. Analogously, it can. We shall argue for this.

The formulation of the second question, or set of questions, Q2 purposely concedes the point made by the likes of Fodor and Piccinini to the effect that there is a theoretical difference between what is called 'functionalism' and what is called 'computationalism', and that though the two can merge (and have in fact merged) so as to form what is known as 'computational functionalism' (for the loci classici of computational functionalism, see Putnam, 1960, 1967b, 1967a; Fodor, 1965, 1968a, 1968b) no such merge is a necessary consequence or requirement of either functionalism or computationalism (Fodor, 2000; Maley and Piccinini, 2013; Piccinini, 2004a, 2004b, 2009, 2010b).[95] We will follow Piccinini and differentiate all three research programs from one another accordingly. Our goal in this section will be to highlight that ultimately, for our purposes, *it does not matter*. When it comes to the application of formal rules and thus to the nature of formal operations (that is, to what it is to apply a formal rule and carry out a formal operation or computation) all three research programs fall prey to Kripke's (or a Kripke-style) argument.

Let's start with functionalism. When talking about functionalism, one would do well to speak of 'the varieties of functionalism' rather than 'functionalism' *tout court*. Bet let us not over complicate things and let us simply attempt to give an overview of functionalism that is at the upmost general level so as to capture the core principles of the functionalist program with respect to the cogitative (specifically, with respect to 'human computation'). Let's start with physical computing mechanisms—physical mechanisms that are engineered or said to compute mathematical and logical functions. Suppose we have three physical computing mechanisms Π_1, Π_2, and Π_3. Suppose all three different devices were designed to compute the addition function and to carry out the same formal operation (say, using the same *program* and *algorithm* for calculating the values of the same mathematical function). Now, Π_1, Π_2, and Π_3 are not physically identical or isomorphic computing mechanisms. In a radically simplified way of talking, we can say that Π_1 is made of transistors, Π_2 of vacuum tubes, and Π_3 of the sort of materials the famous Analytical Engine would've been made of.[96] Of course, their performance will not be equivalent or even approximately similar (e.g., Π_1 will be faster, more reliable, and capable computationally than Π_2 and Π_3, and Π_2 in turn would be all those things with respect to Π_3). The point is that Π_1, Π_2, and Π_3 are all physical mechanisms computing the very same function, even running the same program and carrying out the same formal operations albeit at different speeds, with different memory capacities, and different physical architectures and components. The idea then is that what these physical mechanisms *do* at the formal, rather than at the physical, level is not a matter of types of matter. There is, let us say, physical asymmetry but formal symmetry between Π_1, Π_2, and Π_3. When considering the program that each device runs, or the function each one computes, we need to consider it at a 'higher level', at the abstract level of

---

[95] "Particularly striking in retrospect was the widespread failure to distinguish the computational program in psychology from the functionalist program in metaphysics… (For an instance where the two are run together, see Fodor 1968[b])." (Fodor, 2000, p. 105 n. 4)

[96] This last example is just for my own amusement and is taken form (Heil, 2013, pp. 87–104). Many other kinds of physical computing mechanism will do. It is no mystery that many physical mechanisms today, though physically and architecturally diverse or divergent, are nevertheless taken to carry out the very same formal operations (say, same programming languages and same algorithms, different programming languages but same algorithms, different algorithms but computing the same functions, and so forth). If you like, replace all of Π_1, Π_2, and Π_3 (or at least the last two) with some of the latest physical computing mechanisms available that differ physically yet carry out the same programs, or different programs but same algorithms, or at least compute the very same functions but with different algorithms. Doesn't matter. Our Kripkean argument will cut through regardless of which option you choose.

computation. The formal component (say, the application of formal rules for the computation of a mathematical function) is, physically, multiply realizable and therefore cannot be identified with the physical properties of the mechanism. [97] As Heil rightly notes, the claim is not that computational operations are identical with physical processes (what an identity theorist would claim) but rather that computational operations are *realized* or *implemented*, multiply and variously, by physical processes.[98] Now extend that thought beyond artificial physical computing mechanisms to, say, biological organisms or components of biological organisms (e.g., the brain) and from formal operations to, say, conscious states and psychological phenomena, and you get, with a tweak here and there, functionalism.

The question we need to ask the functionalist as well as the computationalist at this point, particularly when each comes to the study and the explanation, whether in part or *in toto*, of the nature of formal understanding and activity, is this, 'Can definite formal rules and operations be carried out in an intrinsic, mind-independent, and objective way by a physical mechanism (of the right sort) in virtue of its physical properties alone?'[99] If she says yes, she is my enemy and the target of the present philosophical attack. If she says no, then she may or may not be my enemy—we shall see an example of an enemy who answers 'no' to the former question when we consider Dennett (see chapters 4 and 5).[100] To a first approximation, let's say that functionalism is the *metaphysical*

---

[97] To be sure this last statement is a point of contention for *occupant* functionalists, who identify a mental state with the physical processes that satisfy the formal, abstract functional description/specification in question. We will ignore them for the moment, but it should be noted that such formal, abstract specifications are pure formal structures of the sort the physical cannot determinately 'occupy' so it will be subject to our indeterminacy argument. In this regard, occupant functionalism about formal operations will equally falter.

[98] Multiple realization arguments in favor of functionalism trade on that very distinction. The multiple-realizability arguments (MRA) have been one of the most influential arguments in the abandonment of type-identity and behaviorist metaphysical theories and the predominant force behind the turn towards non-reductive physicalism in the philosophy of mind since the 1970s. The general idea is that mental phenomena are multiply realizable, that is, can be realized by different physical phenomena (e.g., properties, processes, mechanisms) and thus do not require a unique *type* of physical properties for their realization. Again, think of a mousetrap. A mousetrap is multiply realizable because it can be realized in multiple devices with different types of physical properties (e.g., wood, plastic, metal, etc.). Thus, so MRA goes, mental and psychological phenomena cannot be identified with any type of physical phenomena but rather must be identified *with that which is multiply realizable*. Yet that which is multiply realizable is multiply realizable in/by the physical. Hence, token-physicalism (as opposed to type-physicalism) generally comes hand-hand with functionalist and computationalist accounts. Functionalist took that which is multiply realizable to be abstract, i.e. abstract functions embedded in a complex abstract functional network, and computational functionalists took that network to be fully computational. Functionalists who did not go along the computational route offered analytic, or causal-theoretic, or teleological accounts of the functional roles of psychological phenomena. But most functionalist accounts identified psychological phenomena with either abstract functional phenomena or with any physical processes that satisfied the requisite abstract functional specifications (see n. 111). The identification of mental or psychological phenomena with abstract structures is what I call 'the formalization' of non-formal (e.g. psychological and physical) phenomena like pain, fear, desires, and like neural activity, electrical circuits, etc. Two further points of clarification. First, MRA does not entail functionalism, and functionalism need not entail MRA. But they are often conjoined, and MRA has been taken as giving strong support to functionalism and functionalism as best accounting for the results of MRA in regard to mental phenomena. Second, MRA, like functionalism, does not necessarily imply token-physicalism. The realization need not be physical. It can *in principle* be non-physical, but this is in general dismissed as philosophically and scientifically problematic and not true in the case of human mental or psychological phenomena, which are necessarily embodied. For more on MRA, see e.g. (Jaworski, 2011, pp. 131–136; Kim, 2011, pp. 129–136; Bickle, 2016; Shapiro, 2000, 2008; Polger, 2015).

[99] A variant, more general formulation of this question can be posed to the physicist and the engineer, 'Can pure formal structures (i.e. mathematical models and equations) be present and instantiated determinately in an intrinsic, mind-independent, and thus objective way by physical properties alone?'

[100] This talk of 'enemy' is no more than a rhetorical device to bring some colourfulness to an otherwise dry exposition of a dry topic. To twist the old adagio, 'the enemy of my enemy is not necessarily my friend.'

theory of the mind that states that what makes something a particular type of *mental* or *psychological* state is dependent upon what it does (i.e., the function or role it fulfils) at a sufficiently high level of abstraction in a given functionally organized system, rather than upon what it (or the system) is made of. Hence the slogan 'matter doesn't matter' for the individuation and specification of a type of mental phenomenon. Types of mental phenomena are said to be *realized* or *implemented* by physical phenomena. Matter matters when it comes to the instantiation (whether conceived in terms of realization, implementation, supervenience, etc.) of types of mental phenomena. Beyond this point, there is very little consensus amongst functionalists.

The varieties of functionalist approaches can be situated along three main parameters or problems (see Van Gulick, 2009, pp. 131–138):

(1) *The extensional problem*. To which mental states or kinds does the notion of functional role apply?
(2) *The definitional/intensional problem*. What counts as a functional role in the relevant sense?[101]
(3) *The conceptual-gap problem*. The problem of relating (1) and (2).

As interesting as this is, it is of no immediate relevance to us. Consider that we are interested in (the capacity for) the ability to comprehend and follow the formal rules constitutive of logic and mathematics. The functionalist about formal operations has in general two main alternatives: either formal understanding and operations, concretely speaking, are the physical instantiation (i.e. realization or implementation) in an individual of the abstract formal rules and formal operations (with which that type of understanding and operation ought to be identified) and thus determinately made present by the physical processes that realize such rules and operations in the individual, or formal understanding and operations are to be identified with the sort of physical processes that fulfil, occupy, or satisfy the given pure abstract formal specifications of the formal rule/operation in the individual who possesses such understanding and carries out such operations. These two

[101] Note that the notion of 'function' at play in this general description of functionalism is not that of mathematical function. This is not to say that it is incompatible with it or that it doesn't encompass it; it is simply to say that it is not presupposed nor identified with it. It refers more to the function a mechanism, or subcomponents of a mechanism, or a set of states, or what have you, play or fulfill in a given functional network. So, for example, the notion of function in the functionalist's jargon is, generally speaking, more akin to the concept of function as used in 'the function the heart plays in the human cardiovascular system'. Again, we are not interested in how functionalism fares with respect to the concept of function thus used except when it is applied to our capacity to *understand* and *carry out* (e.g. compute, define, regiment, etc.) mathematical and logical functions and rules. When it comes to mathematical and logical understanding, or to the *capacity* for understanding and doing mathematics and logic, functional analysis (see e.g., Cummins, 1975, 1983, 2002) alone cannot make intelligible a complete explanation of our capacity for formal understanding and activity since no material sub-components or sub-capacities of the human being can be formally determinate, and the conglomeration of formally indeterminate parts does not generate a formally determinate whole—at least not on account of the mere sum of the parts. The brain, for example, has material (e.g., neurobiological, cardiovascular) functions, not mathematical or logical functions. Thus, functional analysis of the brain will be an analysis within the ontological category of material structures (see chapter 1). To state otherwise is a category mistake which leads to absurdity, as we will show below in the main text and in the following chapters. Yet claims to the effect that the functions performed by a physical system according to its functional analysis amount to mathematical functions computed by the system such that the functional behavior of the physical system can be computationally accounted for has been commonplace (see e.g., Fodor, 1968a; Putnam, 1967b; Cummins, 1975, 1977, 1983; Haugeland, 1978; Marr, 1982; Churchland and Sejnowski, 1999; Eliasmith, 2003). Our argument will show that this cannot be the case. Matter at any level cannot be formally determinate and therefore equating material functions with, or taking material functions to realize determinately, mathematical functions cannot be correct. This is not to say that our brains are not a necessary condition for the acquisition of formal understanding and for its exercise. Remember: we are interested in *sufficient* conditions. No level of material complexity can ever suffice. But there can, of course, be functional analyses and mechanistic explanations of the necessary material (i.e. neurobiological) conditions for *any* kind of human understanding, including mathematical and logical.

responses split up the camp respectively (though in a somewhat confusing manner) between "role functionalists" and "realizer functionalists" (McLaughlin, 2006; cf. Levin, 2018). The former kind of functionalists identify the mental operation with the higher-level, abstract function which is or can be realized in *concrete* operations physically in multiply different architectures or material structures; the latter kind of functionalists identify the mental operation with whichever lower-level physical processes and mechanisms (e.g., the brain) satisfy the higher-level, abstract function or functional specification. Either way, the functionalist will be in trouble. For now, let me note that it is a virtue of Kripke's argument that it does not depend on specific differences between particular functionalist theories but rather on the core and basic principle of the functionalist program: that there are abstract functions or functional specifications (i.e. pure formal structures, cf. chapter 1) that physical phenomena can determinately and objectively, on account of their physical functional or structural properties alone, be an instance of (whether by realization, occupation, supervenience, or what have you).[102]

Though we will focus predominantly on computational accounts since these are taken to offer the best account of formal understanding and our capacity for mathematics and logic, our discussion will illuminate as we go the way in which other versions of functionalism may be refuted if they trade on the distinction between abstract, formal structures (pure formal structures, as we like to call them) on the one hand and physical structures or mechanisms that realize or occupy *determinately and objectively*, in multiple physical ways, the former kind of structures. At this point we break with Kripke, but only partially: our argument is only tangentially an argument against functionalism. It is, first and foremost, an argument against the intelligibility and conceptual adequacy of any materialist explanatory framework, be it reductive or non-reductive, of human computation (of 'the cogitative', i.e. the capacity for formal understanding and formal understanding and activity as such). Functionalism and computationalism are interesting because they trade explicitly on a requirement that the physical cannot meet, namely objective and intrinsic formal (say, mathematical) determinacy. Thus, on our version of Kripke's argument, functionalism and computationalism are *means* to a target, rather than the target itself.[103]

In the case of perception, subjectivity, phenomenal consciousness, psychological states, and so forth, we hint at rather than argue for the abandonment of said distinction for the purposes of constitutive explanations since it conflates a formalization (a kind of abstraction and idealization) of a non-formal phenomenon with the phenomenon itself.[104] In the case of our cogitative powers and activities, we argue (chapters 4 and 5) as follows: it is in the realm of the cogitative powers and acts

---

[102] By objective, I here mean non-dependent on human interpretation, observation, and convention. More generally, non-human or -mind dependent. There is another sense in which a thing can be 'objective' yet be human or mind dependent (e.g., a university or the Supreme Court). Unless otherwise stated, we will mean the former by 'objective'. We will explore something closer the latter use of 'objective' at the end of chapter 3, §7, when talking about the sense of objectivity proper to computational descriptions of and the computations carried out by physical computing mechanisms.

[103] One can, after all, be a functionalist or a computationalist (or a blend of both) about formal understanding while being some sort of dualist. Yet functionalist and computationalist accounts are generally accepted because they provide one of the best explanatory frameworks for explaining naturalistically our cogitative powers. As things stand, they *both* come hand in hand with some version of (token) physicalism.

[104] So we appoint out, though briefly and roughly, that if Ramseification is to have any use in the study of consciousness (i.e. sensation, feelings, perception, emotions, etc.), it must be taken as a formal idealization (a type of formalization) of a non-formal (or non-abstract) phenomenon. As such, what is Ramseified cannot be identified with its Ramseification (i.e. "its" Ramsey sentence). Otherwise indeterminacy issues about consciousness will abound, in which case there would be no fact of the matter as to whether there was such a thing as consciousness and conscious beings, which is absurd. The Ramsey sentence of a mental state must be seen as a formal idealization, with all the formal indeterminacy that idealizations subject the idealized to. Ramseification, however, is *something we (can) do* and an explanation of what that activity and ability amounts to must be given.

of human beings that formally determinate operations and specifications (e.g., mathematized idealizations of physical phenomena) are found. Furthermore, with respect to formal operations, the structure of such operations and intellective acts must be formal. But since material structures can only be said to be in an *extrinsic* and *observer-dependent* way instances of a formally determinate structure, to explain formal understanding as the material implementation of formal structures leads to absurdity (see chapter 4).[105] Yet the distinction, fundamental to functionalism and computationalism (cf. Block, 1980b), between the formal and the material is in itself a wholly appropriate and intelligible one. An abandonment of *this* distinction (between what is formal and what is material and the observation that the latter only approximates or imperfectly realizes the former) is not possible on grounds of intelligibility. Moreover, as far as explaining the nature of formal understanding goes, such an abandonment is incoherent since amongst our cogitative powers and activities lie the ability to think in formally determinate ways, including the ability to formalize, abstract, and idealize mathematically a whole range of phenomena, including physical phenomena. This ability must be explained. When it comes to mathematical and logical functions, both as the object and the form of mathematical and logical understanding and activity, this distinction cannot be abandoned, which is why the problem of bridging the two kinds of structures in an objective and mind-independent way (see, for example, Chalmers, 2012, 2011, 1996) emerges and endures. What must be abandoned is the claim that the gap between formal structures and material structures can be *physically bridged in an intrinsic, determinate, and objective (mind-independent) way*.

Enter computationalism. In general, it provides the best model for explaining the physical implementation of mathematical or logical operations, usually in the way physical computers are said to compute mathematical and logical functions. But differences between functionalism and computationalism must be recognized.[106] First, computationalism and functionalism have each a different origin and came about for completely different purposes.[107] Secondly, they are different in their commitments. Functionalism is a metaphysical theory of the mind. It identifies the nature of mental states functionally on the basis of the role a given internal state or process plays with respect to input states, output states, and other internal states in a given network, considered at a sufficient level of abstraction. It remains neutral as to whether this function or functional role is computational, or causal-theoretic, or teleological, or what have you. But it is committed to the claim that it is physical properties, processes, and mechanisms that realize or satisfy such abstract functions and functional roles. This is why functionalism is considered a version of non-reductive physicalism.[108] An often-cited example is that of a mouse trap, which can be physically implemented in a variety of differently constituted physical devices (e.g. wooden mouse traps, metal mouse traps, etc.), all of which nevertheless count equally well as mouse traps on account of satisfying or implementing the same abstract function or functional role. Such functions or functional

---

[105] It cuts both ways: formal models, simulations, and descriptions of material structures have features the material structures cannot have and the material structures have features their formal models, simulations, and descriptions cannot have.

[106] It must also be recognized that computational functionalism has not been met with open arms. For representative critiques of computational functionalism, see (Putnam, 1988; Searle, 1992; Lucas, 1996; Penrose, 2011b)

[107] For an overview of the different historical and intellectual origins of both functionalism and computationalism, see (Piccinini, 2004a, 2004b, 2009, 2010, p. 269)

[108] Functionalism does not requite token-physicalism. One can be a functionalist and dualist. But historically the motivating factor behing functionalism in the philosophy of mind was that it was able to overcome the pitfalls of logical behaviorism and reductive (i.e. type-identity) physicalist theories without renouncing to some form of physicalism in metaphysics. Hence, though a mouse trap is an abtract functional system, only physical mechanisms can actually realize and count as mousetraps and only in virtue of playing the apropriate functional role (i.e. trapping mice) rather than in virtue of any specific or type of physical property (e.g. there's no unique material configuration for mousetraps; they can be implemented in a variety of differently constructed physical mechanisms).

descriptions can but need not be computational, and in fact most forms of functionalism today are not computationalist.

Computationalism on the other hand need not be functionalist in the sense that it need not reduce the nature of mental states or powers to functions or their functional characterization, and thus need not be committed to the claim that the nature of mental states is ultimately functional or that the mind is but the physical realization of abstract functional descriptions. This is a little harder to spell out because computationalism *is* the view that "the functional organization of the brain (or of any other equivalent system) is computational, or that neural states are computational states." (Piccinini, 2010a, p. 271) But the idea is that computationalism is compatible with a non-functionalist metaphysics of the mind. Now, there are different versions of computationalism depending on how the notion of 'computation' is understood.[109] Computationalism and functionalism, however, both agree that abstract structures whether considered merely functionally or merely computationally or both can be realized or implemented in multiple physical ways and by diverse physical mechanisms. This is a minimal requirement of both views, and a necessary one of any view that purports to explain the computation or application of mathematical functions and rules *in terms of* physical capacities, mechanisms, and processes. Alternatively, it is a minimal requirement of any view that claims that physical computing mechanisms can compute mathematical and logical functions on account of their intrinsic, causal-physical, properties and structure alone. This gives us quite a bit of leeway to use 'computational functionalism' (functionalism, for short) and 'computationalism' interchangeably in developing Kripke's argument since the argument does not depend on any difference between the two research projects but rather on a fundamental presupposition common to both: that the physical can realize or implement (concretely compute) determinately and objectively on account of physical properties alone mathematical and logical functions and operations.

So we concede that computationalism and functionalism are not necessarily the same in the sense that if computationalism is true, it does not follow that functionalism is true, nor in the sense that if functionalism is true, then computationalism has to be true (Piccinini, 2009, p. 515). A variety of non-computational functionalisms exist (see, e.g., van Gulick, 2009, pp. 131–138) and the same seems to be the case on the other end, non-functionalist computationalism (Piccinini, 2004b, 2004b, 2004a, 2008a, 2009, 2010a, 2010b, 2015; Piccinini and Scarantino, 2011). But if the mathematical and the logical is to be implemented, realized, or occupied by anything physical in a determinate and objective way, there must be some fact (that is, some physical fact, structure, or mechanisms) that determines which mathematical (or logical) rule or computation is being implemented. If there is no such thing as the objective physical determination of formal operations and structures, both functionalism and computationalism will stand refuted as an explanatory model of the objective capacity and ability of human beings to do mathematics and logic (or more narrowly, for mathematical and logical computation). Again, it is a strength of Kripke's argument that it neither rises nor falls on what is different but rather on what is shared or common.

It is perhaps telling that Kripke's argument against computationalism seems to have emerged as a consequence of entertaining the possibility of computationalist or functionalist responses to the skeptical paradox.

---

[109] "Realization, then, is a relation between abstract descriptions and concrete systems whose states match or correspond to the states [or functions] postulated by those descriptions." (Jaworski, 2011, p. 140; cf. Kim, 2011, pp. 129–167, 169–191) We will have time to consider, at least in the case of computationalism, accounts of realization other than this one, which have come to be known as 'simple mapping accounts' (Godfrey-Smith, 2009) and 'structuralist accounts' (Rescorla, 2013) of realization or concrete computation.

> Especially I have feared that some readers of the discussion in the text will think that 'functionalism' is precisely the way to modify the crude dispositional theory so as to meet the criticisms (especially those that rely on the circularity of *ceteris paribus* clauses). (I report, however, that thus far 1 have not run into such reactions in practice.) I cannot discuss functionalism at length here without straying from the main point. But I offer a brief hint. Functionalists are fond of comparing psychological states to the *abstract* states of a (Turing) machine [or of a causal mechanism], though some are cognizant of certain limitations of the comparison. All regard psychology as given by a set of causal connections, analogous to the *causal* operation of a machine. But then the remarks of the text stand here as well: any concrete physical object can be viewed as an imperfect realization of many machine programs. Taking a human organism as a concrete object, what is to tell us *which* program he should *be* regarded as instantiating? In particular, does he compute 'plus' or 'quus'? If the remarks on machines in my own (and Wittgenstein's) text are understood, I think it will emerge that as far as the present problem is concerned, Wittgenstein would regard his remarks on machines as applicable to 'functionalism' as well.
>
> (Kripke, 1982, p. 37 n.24)

Kripke has in mind the good old computational functionalist approaches of Putnam and Fodor. But in his lectures against computationalism, he believed that his argument worked against *any* variety functionalism, even non-computational ones, though he formulated the argument only in terms of its computational versions. "Though sometimes people distinguish between functionalism and machine [computational] functionalism, I think that the type of remarks I make, if valid, would be valid for any [functionalist] view, including one that interprets psychological statements about me as statements on a high-level of abstraction about my causal hookup." (Kripke, 1984, p. 4) We will not pursue this more generalized line of reasoning, although much will be said now and there that will point to a nod of agreement from our part.

Another interesting feature of Kripke's argument against functionalism and computational theories of the mind is that he found that they failed were they were (and still are) generally taken to succeed: explaining our higher cognitive capacities and abilities (our cogitative powers) and their unique formal and conceptual features (features such as Fodor's triplet of productivity, systematicity, and inferential coherence. See e.g. Fodor, 1985, 1989; Fodor and Pylyshyn, 1988; cf. Fodor, 1975, 2008, 2000). Much ink has been spent trying to show that functionalism/computationalism cannot capture, in any explanatorily adequate way, the most fundamental aspects of consciousness, of subjectivity, of personal identity, and so forth, and is therefore defective as a holistic theory of the mind. But, comparatively, very little ink has been spent in showing that it cannot account for logical and mathematical understanding and activity, and the ink that was spent traded on the philosophical significance of certain mathematical theorems and proofs (e.g., Gödel's incompleteness theorems and Turing's proof of the undecidability of the halting problem) for showing that the human mind was not (or was more than) a computing machine. Yet these mathematically inspired approaches did not question (in fact, conceded) that physical mechanisms could do mathematics (to be precise, could compute mathematical and logical functions) and be computing mechanisms; the problem was that we could do something that no computing machine could and that as a consequence *we* could not be computing machines, physical or otherwise (Lucas, 1996, 1984, 1996; Penrose, 1996a, 1999, 1996c, 2011b). Kripke's argument is much more fundamental than these Gödelian arguments and is on a whole different playing field. It has more to do with the nature of physical computing mechanisms than anything else. For, so long as any theory, whether computationalist or functionalist

or what have you, holds the formal structure and features of mathematical and logical operations are but the physical realization or implementation of (algorithms for computing) mathematical logical functions, they will fail. No physical property, no physical computing mechanism, can determinately and objectively realize or implement a *definite* formal operation on account of its physical properties alone. Once one begins to consider the operations at the appropriate level of abstraction in which they count as formal (i.e., mathematical or logical), the physical mechanisms that supposedly implement a given definite pure function (say, the addition function) will equally well count as implementing an incompossible definite pure function (say, the quaddition function). This is a claim entirely different from, though not incompatible with, the anti-computationalist claims made on the basis of Gödels disjunction (for an excellent treatment of the arguments and the debate from Gödel to Lucas to Penrose, see Koellner, 2018a, 2018b; for a recent set of essays on Gödel's disjunction, including a paper by Koellner, see Horsten and Welch, 2016).

As we will see, a consequence of Kripke's anti-functionalist argument is that formal determinacy in the case of physical computation is ultimately fixed by the intentions of the designer (and in the case of non-artefactual mechanisms like the human brain if taken be computational, by the relevant observers-interpreters).[110] All of this requires practices in which there's agreement in assigning physical items the status of symbols, of symbol and of symbol manipulation so as to be able to pick out which physical processes and mechanisms involved can count as relevant for the computation of a *determinate* pure function. This is so for artificial as well as for organic 'computing' mechanisms. Any neural correlates to a given formal activity will by themselves satisfy incompossible formal descriptions. These are just not the sort of things that could determine a definite formal rule (say, addition as opposed to quaddition). Physical processes and mechanisms of whatever kind can always be interpreted differently with respect to the *form* of the computation because they cannot in and of themselves exclude incompossible formal descriptions and operations from being satisfied at the same time and in the same respect. This is alright for the physical, and for practical purposes formal indeterminacy neither hinders nor disturbs the engineering of physical mechanisms for computational purposes nor the study of physical phenomena in general, organic or otherwise; but this has preposterous and absurd consequences for formal understanding and intellectual activity as a whole (we will show this in chapter 4) if that sort of activity is but the physical realization or implementation of formal structures.

We can call functionalism 'realization physicalism'[111] (Jaworski, 2011, pp. 144–149; cf. Block, 1980b), though this is misleading (see our distinction above between role and realizer functionalism),

---

[110] There's a metaphysics that considers all things artefacts (e.g. Dennett, 1991, 1987). We will come back to it in Chapter 4. To be clear, Kripke's argument is *not* strictly speaking a triviality or a liberality argument against computational functionalism (see Godfrey-Smith, 2009). Triviality arguments (Putnam, 1988; Searle, 1992) and liberality arguments (Block, 1980b) are closely related but are somewhat different: the former argue that anything can be made to count as a computing mechanism and therefore that anything can be said to compute, even rocks, whereas the latter purport to show that at least some things which should not count as computing or as having certain functional properties end up counting as having them under proposed functionalist or computationalist theories. The conclusion of Kripke's argument is rather different: that no physical computing mechanism can follow *any* formal rule, for it cannot follow any definite formal rule.

[111] This wording may cause some confusion, as some characterize what we call 'realization physicalism' as 'role functionalism', letting 'realizer functionalism' signify instead a philosophical account that does not take pscyhological phenomena to be abstract. Oversimplifying, on this distinction 'realizer functionalism' refers to those functionalist theories which do not view psychological and mental states and powers as identifiable with higher-order or higher-level abstract functions or functional role but rather as identifiable with the *lower-level* properties that satisfy the relevant abstract functional specifications. Thus, realizer functionalism considers such higher-level phenomena as mere abstract functional characterizations or specifications of the lower-level physical processes or objects that actually fulfill or constitute the psychological roles or functions in question. Realizer functionalism differs from 'role functionalism' in that the latter *does* take the relevant psychological and mental states and powers to be identifiable with higher-level or

since it takes mental phenomena to be (i) a species of abstract phenomena (i.e. the abstract objects, properties, states or relations postulated by an abstract theory) that (ii) are realized by physical phenomena (i.e. the concrete, physical objects, states, properties and relations that are said to realize, satisfy, or correspond to, the inventory of abstract objects, properties, etc., postulated by the abstract theory).[112] The computational approach of functionalism is but one way in which (i) and (ii) can be cashed out (Ramseification being another way). Buechner illustrates this clearly when detailing why Kripke's argument applies equally well to non-computational functionalist approaches:

> The causal-theoretical functionalist does not conceive of the human mind as a computing machine, but rather as instantiating the states of commonsense "folk" psychology. Whatever level of abstraction causal-theoretical functionalists choose for their total psychological theory, they must employ the distinction between the abstract states of an abstract causal network and the purely physical states of the brain realizing that causal network. But that *is* the basic distinction that gives rise to Kripke's problem for functionalism. It does not matter whether the abstract causal connections in the abstract causal network constituting the total psychological theory of human beings can or cannot be modeled in a computational structure. Kripke's refutation of functionalism demands only that the functionalist posits a distinction between an abstract diagram and the physical object physically realizing it. Causal-theoretical and machine functionalists make that distinction. That is why Kripke's refutation of functionalism targets them both.

---

higher-order functions or functional roles and not with any lower-level properties. Lower-level properties are generally take to be physical, though functionalism does not necessitate this. According to role functionalism, lower-level physical properties, though not identifiable with the higher-level properties that constitute psychological states or powers, are and can nevertheless realize such abstract, higher-level properties in numerous ways (see e.g. Levin, 2018). Our Kripkean argument is directed manly against any version of 'role functionalism', at least as it concerns formal understanding and activity. But it applies to 'realizer' functionalism just as much. Any physical property or process that can be said to satisfy a certain abstract (e.g. formal) specification can be equally well said to satisfy some incompossible abstract specification. Now, the realizer functionalist might take this as a point in her favor, since her claim is precisely that since psychological states cannot be identified with abstract functions or abstract functional roles, psychology must be a matter of lower-level physical properties. But since any lower-level property or process, if wholly physical, will satisfy incompossible abstract functional specifications or descriptions, it is hard to see how any functional specification could be determinately and objectively true of a physical property or process. At any rate, when it comes to mathematical and logical functions, it makes no sense to identify these with lower-level *physical* properties or processes.

[112] We contrast "realization physicalism" with "supervenience physicalism"—the second most popular form of non-reductive physicalism in contemporary philosophy of mind (see e.g. Kim, 2011, pp. 8–11; Leuenberger, 2008; McLaughlin et al., 2009; McLaughlin and Cohen, 2007). Now, if the notion of *supervenience* is built into the explanatory framework for explicating how the physical can instantiate, satisfy, exemplify, or carry out pure formal structures, as supervenience physicalism may surely attempt to do, then it is effectively refuted by the Kripkean argument. There are many ways of cashing out the notion of supervenience, but in general we say that a set of properties A supervene on a set of properties B just in case there *cannot* be two things with exactly the same set of B properties that differ with respect to their A properties. That is, there can be no two things which differ in their A properties but do not differ in their B properties. The idea is that the supervenient set of B properties determines a unique supervening set of properties, namely the A properties, and therefore anything with B properties will necessarily have the same A properties. Thus, A supervenes on B. "No change in A without a change in B", is the slogan. So the notion of supervenience would presumably require that any physical property and process that realizes an abstract function must make it so that the latter supervene on the former such that the physical properties and processes in question necessarily determine a unique set of abstract functional properties. But Kripke's argument entails that no fixed set of physical properties can determine a unique pure formal structure, thus showing that with respect to pure functions the physical violates supervenience. So if I am correct, the physical will necessarily violate supervenience with respect to formal structures, though it can nonetheless be said to realize formal structures in an *extrinsic* and *observer-dependent manner* (we show this later in this chapter; cf. §7 chapter 3).

(Buechner, 2011, p. 348)[113]

Generally speaking, functionalists clash with each other on three fronts: 1) on how to specify the abstract theory, 2) on what precisely the ontological 'relation' between psychological states and the objects postulated by the abstract theory is, and 3) on what it means to physically 'realize' such, abstractly considered and specified, states.[114] As we've said, we can leave functionalists to sort these issues among themselves. So long as the functionalist draw the distinction that is necessary and gives rise to the very functionalist project, the distinction between the abstract function or specification of a function and the array of concrete *physical* systems that multiply realize such an abstract function,

---

[113] Here's a simple sketch of the way in which Kripke's problem can be formulated for causal-theoretic functionalism through the Ramsey-Lewis method. Consider the following "Ramseified" theory of pain (see Kim, 2011, p. 169-191):

> (T_R) There exist states M1, M2, and M3 such that for any x, if x suffers tissue damage and is in M1, x is in M2; if x is awake, x tends to be in M1; if x is in M2, x winces and groans and goes into M3; an "if x is either not in M1 or is in M3, x tends to make more typing errors.

And its abbreviation

> (T_R) *x* is in pain $=_{def}$ ∃M1, M2, M3[T(M1, M2, M3) and x is in M2]

Notice that this is a "topic-neutral" definition (Smart, 1959). The lettes $M_1,\ldots, M_n$ are predicate variables (just like *x*s and *y*s in mathematics are "individual variables") that are taken to be topic-neutral logical terms, neither psychological nor physical. Expressions like "is in pain", "is normally alert", "is distressed," and so foth are taken to be predicate constants, that is, actual predicates. Now, if one is a functionalist, one will say that there is a physical property, or a set of physical properties, of whatever kind that can, and do, realize (T_R). The problem is, in part, that (T_R) is an abstract, i.e. formal, description and definition of pain. As such, it will be imperfectly realized by anything physical, and will be subject to incompossible formal definitions. But nobody experiences pain imperfectly—by which I mean that whenever someone is in pain, is not by approximation to (or simulation of) pain. One can always construct a formal definition incompossible with (T_R) (say, "quain", which yields exactly the same input-states-output patterns and relations for the first fifty-seven occasions but yields an entirely different mental state thereafter) that the purported physical realizer of (T_R) equally well can said to realize. Take the following incompossible definition,

> (T_{R^*}) *x* is in quain $=_{def}$ ∃M1, M2, M3[T(M1, M2, M3) and *x* is in M2 if less than the 57th time; otherwise, M4]

where 'M4' is an unpredictable state. One can get more complex and creative, but this will do. Anything physical that satisfies (T_R), satisfies (T_{R^*}). But since this violates (DR), we have that no physical thing can determminately satisfy a unique mental state, and thus it realizes determinately none. Hence, no material being (including animals) can be objectively or observer-independently in pain. But this is absurd. Note also that the phenomenal aspects of the experience in question (including what analytic philosophers call the *qualitative*, the subjective feel, i.e. "what it is like", of experiences) will be equally compatible with both (T_R) and (T_{R^*}). In short, the conclusion should be not that pain is an immaterial experience but rather that pain is not an asbtract (i.e. formal) function and that being in pain cannot consist in the physical realization of an abstract functional role. In other words, pain (i.e. "hurting") is *not* a pure formal structure.

[114] Unlike Kripke, I do not think the argument applies to *any* functionalist theory of the human mind. But it will apply, however, to any functionalist theory of the mind that identifies psychological states or experiences with the abstract functions of a complex abstract causal network that is realized by a physical thing. But to identify psychological states with abstract functions is to convert them all into pure formal structures, something which psychological states like "being worried" are surely not. Ramsey sentences, for example, *formalize* psychological states and experiences, but functionalists ignore this and proceed, confusedly, to identify the psychological state or the experience with its formalization or with the physical processes that realize or satisfy the formalization. But the formalization is a human abstraction and idealization, a pure formal structure, which cannot uniquely and determinately be realized by anything physical. Plus, psychological (i.e., qualitative-phenomenal) phenomena, like physical phenomena, are equally formally indeterminte. Only human understanding and activity can be said to be formally determinate, and thus to be intrinsically and objectively a case of a pure form. We will have more to say about this in chapter 4.

she is in trouble. At any rate, in the case of pure functions like mathematical and logical functions, the distinction is inevitable.[115] If functionalist physicalists want to explain naturalistically what it is for us to follow pure formal rules, to compute and realize pure functions, then they must account for it in terms of physical (e.g. neurobiological) processes and mechanisms that somehow determine the formal structure and character of our mathematical and logical operations. Some versions of functionalism may eschew talk of abstract functions and abstract descriptions altogether.[116] Fine. They are save (at least from us). Our argument will refute only those theories whose explanatory models trade on the distinction between an abstract function, rule, operation, or description on the one hand and the physical properties, processes, and mechanisms on the other that purportedly realize or implement determinately and objectively the former kind of functions. Explanatory frameworks that do not concern themselves with the cogitative, with formal understanding and activity such as the application of formal rules and calculation of mathematical functions, and which do not end up making or converting (whether advertently or inadvertently) sensory and psychological phenomena into a species of formal phenomena[117], and which neither require nor presuppose a realist view of the physical computation of mathematical and logical functions, will not be refuted by our Kripkean argument. But for any philosophical account concerned with the nature of formal understanding, and therefore with the nature of the array of operations and activities that constitute its exercise (e.g. carrying out definite mathematical operation in accordance with the relevant formal rules), Kripke's argument will be unavoidable. Furthermore, if one is a physicalist of either the type or token sort and believes in the physical realizability and implementability of pure functions, the task of trying to bridge the gap between the two kinds of structures (the formal and the physical) becomes a monumental but necessary task (Chalmers, 2012, 2011, 1996). We show that no bridge is possible.

3. THE ARGUMENT FROM DETERMINATE REALIZABILITY

Let us extend Kripke's original plus/quus considerations to physical computing mechanisms in the relevant and interesting sense, namely physical mechanisms that are taken to compute mathematical and logical functions on account of their physical properties, parts, and configuration alone.[118] Consider the following scenario:

---

[115] It is not clear how non-computational versions of functionalism would account for mathematical and logical reasoning. The idea seems to be that, if nothing else, computational functionalism works at least for that kind of higher order mental activity, but not for any other mental phenomenon (e.g. consciousness or intentionality).

[116] Teleological versions of functionalism, which take function in the rich sense of biological function, do *not* fall prey to our argument so long as they don't cash the biological function in terms of computation in any logically or mathematically interesting sense. Teleological functionalism would be in hot water if it attempted to explain our capacity and ability for formal understanding in terms of biological functions, since the biological is just as formally indeterminate as the chemical, and so forth. Again, this will be briefly touchd upon in the next chapter.

[117] We are of the view that re-describing sensations like pain or hunger in Ramsey-sentences commits precisely this: it *formalizes* a sensation and then identifies the metaphysical nature of the sensation with the formal structure (i.e. the relevant Ramsey sentence). Kripke's argument then can be run with such sensations, e.g. pain and qupain. There is nothing wrong with formalizing sensations, so long as we are not misled into thinking that the nature of a sensation can be identified with its formalization, which is then reified and said to be multiply realizable.

[118] I am assuming, for the sake of argument, that there is a non-trivial characterization of 'physical computing mechanisms' *contra* Putnam and Searle. I side with Piccinini (Piccinini 2015, 2017) that this characterization is not trivial, and that some things just don't count as computing mechanisms. So, again, our Kripkean argument is not one about triviality. It concedes, for the sake of argument, the resolution of triviality problems.

> Suppose we came across a computer which we were prone to describe as being able to add correctly when operating with numbers up to and including 57 but as always going wrong with higher numbers because, with higher numbers, whenever we tried it, it came up with the answer 5. We might be misdescribing it if we did this. It might not be an *adding* machine at all. It might be what, following Kripke, we could call a 'quadding machine'. The quus function, as defined by Kripke, would give a rule which the machine could be obeying throughout. It seems we could say that it might all the time have been a correctly functioning quadding machine. But if it were *always* possible, no matter what "answer" the machine might come up with, to give a rule according to which it was operating "correctly", then, it seems, we could never really make sense of the idea of the machine operating "incorrectly" (or "correctly", for that matter) and therefore no sense of it operating according to a rule at all. There could, therefore, be no sense in describing it as either an adding or a quadding or any kind of computing machine.[119]

This is, in a nutshell, the thrust of Kripke's argument against computational functionalist theories of the mind—and the basis for our claim that Kripke's argument is, ultimately, an argument for the formal indeterminacy of the physical. The problem generalizes:[120] it can likewise be formulated using *any* pure function or rule, not just the 'addition-quaddition' pair of incompossibles. (Pure logical functions, like *modus ponens* or quantifier rules, would do just as well.)

Our version of Kripke's anti-functionalist argument goes as follows:

A) Take any physical object or mechanism whatsoever, call it φ, and constrain it however much you'd like so that it cannot count as computing every function and so that we can label exactly its properties and behavior. (No triviality problems, and no labelling problems).[121]
B) Suppose φ is a physical system whose purely physical changes and processes compute a pure function *f* (say, the addition or the identity function).
C) Then there will be *at least* one other incompossible pure function *g* which all the physical changes, processes, and facts of φ equally well satisfy and sanction.[122]
D) Thus, φ cannot exclude incompossible functions from being equally realized or satisfied.
E) But incompossible pure functions, such as *f* and *g*, by definition exclude one another from being equally satisfied or realized at the same time and in the same respect.
    a. For, in part, a genuine case of a pure function is determinate with respect to which counterfactuals are true of it. But a case that is both genuinely of *f* and genuinely of *g* at the same time and in the same respect will equally satisfy incompossible, i.e. contradictory, counterfactuals—which is absurd.

---

[119] Slightly modified from (Stock, 1988, p. 326)

[120] Note that Kripke's unpublished argument builds upon his plus/quus considerations. After elaborating the "skeptical paradox", he considered a host of possible facts that could be brought in to dissolve or solve the paradox. All the facts considered were (a) physical and material facts (including dispositions to behave in certain ways), (b) historical facts, and (c) phenomenal facts. All of them were found to be indeterminate with respect to the rules of addition and quaddition. Note also that Ramsey sentences in causal-theoretic versions of functionalism are but the translation of psychological phenomena into pure formal structures, but since pure formal structures have no (nor do they determine any) physical and phenomenal properties, Ramsey sentences employed to "capture" the structure of mental states end up making psychological structures a species of formal structures and thus are in principle condemned to fail to accurately understand and explain our psychological life, including its physical and "phenomenal" aspects.

[121] See (Buechner, 2011).

[122] This should be read with the qualification 'at the same time and in the same respect', which will be presupposed throughout.

F) Therefore, since nothing about φ can exclude incompossible pure functions from being equally well realized or satisfied, it follows that no such pure function is and can be realized or satisfied by φ.
G) Therefore, no φ can realize a pure function (i.e. φ is formally indeterminate). [123]

Proposition E) is the logical ground for what we call the principle of 'Determinate Realization', which we adopt but adapt from Lande's *Unique Realization Principle* (Lande, 2011, p. 2).

> *Determinate Realization (DR)* = For any realizer *P*, *P* realizes *f* (where *f* is a variable ranging over pure formal operations) if and only if all the features of *P* relevant for the realization of *f* determine that it realizes *f* and only *f*, such that for every *f** that *P* realizes, *f** = *f*.[124]

Let's be clear on what (DR) is *not* saying. It is not saying that the same set of physical properties or mechanisms relevant for the computation of one function (e.g. addition) cannot also be relevant

---

[123] Even though I am focusing on the *form* of arguments or inferences in a way that makes *it* heuristically separable, and sharply distinct, from the *content* of the arguments or inferences in question, I do not intend in any way to suggest that actual formal reasoning is, as capacity and ability, independent of semantics or conceptual intentionality; likewise, such sharp distinction and contrast as I draw between formal aspects and semantic aspects of language is not meant to suggest that syntax and semantics are independent in natural languages—as if the syntax of natural language were not to some extent determined and shaped by semantics and pragmatics. If you are a Chomskyan, you will claim that it is thus and that syntax and semantics are, in principle, independent of one another; but if you are, say, a Wittgensteinian, you will take such a claim to be conceptually confused. So let me just say this: even if it's granted that the claim that syntax and semantics are (specifiable) independent of one another is a philosopher's and cognitivist linguist's chimera, and in no way elicited by nature of language, there's no doubt that we can construct artificial languages in which such an 'independence' between the two holds (cf. context-free grammars, and Hilbert's notion of the ideal or infinitary sector of mathematics). Even if it's true that the syntax of natural languages cannot be properly characterized and specified independent of semantics and pragmatics, *contra* Chomsky, you cannot reasonably deny that we can construct formal languages in which it can be thus characterized and specified (e.g. the Principles & Parameters theory in generative linguistics in just one example). In other words: I am not, for present purposes, interested in the question whether it is true that the syntax and lexical factors of a natural language are specifiable in formal ways independent of semantics and pragmatics, but on the question what is it that we are doing, i.e. what is the nature of the ability and the activity we are engaged in, when we think in any formally determinate way, including when we think and specify a syntax in formal ways independent of semantics and pragmatics. That's all I need for my argument. I am interested in the capacity and ability to do and think in such terms, and the activity that results upon the exercise of that ability, not on whether natural languages are genuinely Chomskyan or not. Just as this is *not* an essay on the philosophy of mathematics, it is likewise *not* an essay on philosophical grammar, philosophy of language, and philosophy of linguistics. It will, however, have consequences for all such disciplines, as we will see in chapter 6.

[124] This is a variation of Lande's principle of *Unique Realization* (UR) (Lande, 2011, p. 2). Lande formulates the principle specifically with respect to finite state automaton (FSA), though he also claims that it is generalizable. We've formulated the generalized version. In addition, though Lande formulates this criterion more as a dialectical tool, namely as something the functionalist-cum-token-physicalist *must* accept if he wishes to remain one, I take the criterion as a necessary condition for the realization or computation of *any* pure formal structure, and thus as something *any* realizer must satisfy in order to realize a pure function. Thus, as opposed to Lande, I take something like (UR) to be part of the definition of realization,

> A physical system realizes a pure formal operation *f* if and only if it realizes only that *f* and no other.

I will come back to Lande's rather weak objections to this move (cf. Lande, 2011, p. 7). For now, suffice to state that what it is for a definite pure function to be genuinely realized in a given case is, in part, to effectively exclude any other incompossible pure function from being equally well (i.e. at the same time and in the same respect) realized. We have already seen in a preliminary way why this must be so. Hence, if something cannot satisfy this condition, it cannot genuinely be said to realize a pure function—except derivatively and extrinsically (I will come back to this later).

for the computation of another function (e.g. subtraction).[125] It is saying that the set of physical properties in a given mechanism, suitably configured, that are taken to be the ones relevant for and the determinant of a definite operation (e.g. addition) must determine a definite operation (e.g. addition) and not any other incompossible one (e.g. quaddition)—which is to say that it must not be indeterminate with respect to which function it is relevant to and the determinant of. It is not saying that a physical computing mechanism cannot compute multiple different and divergent functions at the same time. It is saying that it cannot compute multiple and divergent functions *at the same time and in the same respect*. If *per impossibile* something could be said to implement incompossible functions at the same time and in the same respect, then there is no such thing as something computing one function rather than another and therefore any function at all. The principle (DR) points precisely to the fact that multiple *divergent* or *incompossible*[126] functions cannot be realized or computed at the same time and in the same respect. Whatever cannot exclude divergent functions from being equally well realized at the same time and in the same respect cannot be said to realize any one of the divergent functions. Whatever in *P* constitutes the *determinant* of a definite function (e.g. the addition function) must guarantee that no other *incompossible* function (e.g. the quaddition function) is in the same respect and at the same time determined. There can of course be simultaneous computations in a given system (e.g. parallel processing and quantum computation), but not in the same respect.[127]

---

[125] There is an insignificant problem with Lande's formulation of the principle in terms of *"all* the facts of the realizer…". For, surely, there are physical facts that *may be* totally irrelevant for the computation of an *f* (say, the color or aesthetic design of *P*). We've corrected this "glitch" in the formulation of the principle.

[126] Perhaps someone would want to say that *different* functions like addition and subtraction are not necessarily *divergent* (say, because there is no point prior to which their computations do not diverge for a shared segment of possible arguments), and reserve 'being incompossible' for the incompatibility and mutual exclusion of *divergent* functions like addition and quaddition, now taken as functions whose values coincide for a shared finite or infinite segment of their possible arguments but diverge for the totality of their possible arguments, like addition and quaddition for numbers greater than 57 but unlike addition and subtraction for any pair of positive integers. But this seems to be quite vacuous. Take addition, subtraction, and multiplication and consider each function over the ordered pair ⟨0, 0⟩; then addition, subtraction, and multiplication will trivially count as divergent functions. For, even though they diverge for any ordered pair of positive integers (e.g. ⟨1, 1⟩), they yield the very same value for at least one pair of natural numbers ⟨0, 0⟩. Now suppose there's a physical computing mechanism that can only carry out computations over the ordered pair ⟨0, 0⟩ and which outputs '0' for all its computations. Then all the computations of the computing mechanism could count at the same time and in the same respect as having been cases, or as being cases, of multiplication, subtraction, and addition; for, multiplication by 0 yields 0, adding 0 to 0 is 0, and subtracting 0 from 0 yields 0 (division being nowhere predicable since division by 0 is an excluded operation, though one can re-interpret the output '0' as signifying 'null' or 'excluded operation'; in which case it will be following the rules of division as well!). But this is absurd. Trivially, all non-equivalent functions (e.g. functions that have a different domain and/or range) are incompossible. But perhaps the idea is to make the point against whoever says that physical computing mechanisms can be structured in such a way as to exclude *different* standard arithmetical functions from being computed at the same time and in the same respect (e.g. addition as opposed to subtraction, its inverse operation) though not as to exclude non-standard *divergent* arithmetical functions (e.g. addition as opposed to quaddition, it's incompossible operation). For, this is simply confused because any standard function has at least one non-standard incompossible function it can be paired with (e.g. subtraction/qubtraction), and in order for any physical mechanism to realize *any* definite pure function, whether standard or non-standard, the relevant divergent-incompossible function(s) must be excluded. Hence, not even standard, i.e. 'different', functions or rules can be determinately computed or applied—since every function has a divergent function. Therefore, physical mechanisms are formally indeterminate. This is what we want to prove. I must confess I find no need for this maneuvre, using 'different' and 'divergent' to name diverse set of functions or rules, although this distinction might come in handy when considering Chomsky's (Chomsky, 1986) critique of Kripke's Wittgenstein on rule-following. But other than that, I think the distinction is more liable than not to cause confusion. So I will not employ it (but see chapter 6).

[127] If, as some would suggest (cf. Ellis 2014), mathematical and logical principles and rules breakdown or don't apply at the fundamental level of reality (e.g. at the quantum level), then this would just entail that the physical at its most fundamental level cannot be either the cause or the constitutive determinant of the formal determinacy found in mathematical and logical activity. A level of reality that cannot, on account of its nature, allow for the exclusion of

Suppose that the computationally relevant states in any concrete computation are and must be wholly physical (Piccinini, 2015; cf. Stabler, 1987, p. 5). If *P* is wholly a physical mechanism, then all the features of *P* computationally or functionally relevant for the realization of *f* will be physical. In this context, (DR) makes it clear that for a formal structure *f* to be realized by any physical realizer *P*, all the physical facts (e.g. input states, transition-causal states, parts, and output states) of *P* that determine *f* must entail that it is *f* and only *f* that it is being realized by *P*. For example, if this calculator in my hand is computing the addition function, then whatever in this physical device determines the computation of the addition function must entail that it computes the addition function and not the quaddition function, or some other incompossible one. That is to say, it must determine a function *to the exclusion of any other incompossible function*.[128] As Lande rightly puts it, the principle "requires that the realization relation must be fully determinate and unique." (Lande, 2011, p. 2)[129] We will later show why anyone who holds to the intelligibility of concrete computation cannot avoid this principle. For now, I just want to press home the point that Kripke's argument provides all the grounds necessary for showing that no physical object and process can satisfy (DR). We will later see where Kripke's skeptic (Kripke 1981, 1982) went wrong; for, from the fact that physical (and phenomenal) processes and features fail to satisfy (DR) it does not follow that therefore our logical and mathematical reasoning fails to satisfy (DR)—unless one also assumes that such reasoning is likewise entirely physically (and phenomenally) constituted and determined. We will come back to this principle later.

---

incompossible functions and rules, and for their non-identity, cannot be the sort a level of reality suitable for the computation and application of definite pure functions and rules. This is actually a point in our favor.

[128] (DR) applies to any potential realizer, whether physical or non-physical.

[129] David Lewis (Lewis, 1972, p. 251ff) (Lewis 1972, p. 251) talks about 'realization' and 'unique realization' in a different way, the latter being included in the meaning of the former but not the other way around. For Lewis, a physical thing *P* uniquely realizes a function or functional role *f* if no other physical thing except *P* satisfies that function or functional role, *f*. Notice that this is quite different from our characterization of 'unique realization'. For us, a physical thing *P* uniquely realizes a function or functional role *f* if no other function or functional role except *f* is 'realized' or satisfied by *P*. (I am taking for granted the "at the same time and in the same respect" qualification.) Lewis is talking about those physical objects of which it is true that they and only they realize the functions or functional roles in question. Thus 'unique realization' concerns those functions or functional roles that are satisfied or realized by one and only one (set of) physical object(s). Our problem is quite in the opposite direction. We are talking primarily about the abstract functions themselves, which require for their *real* realization and satisfaction that no other incompossible function be realized at the same time and in the same respect. Hence, our issue is whether there is or can be any physical object of which it is true that it satisfies or realizes an abstract function to the exclusion of any other incompossible one. This characterization of 'unique realization' with respect to the physical concerns, not whether some function has a unique physical object that satisfies it (that is, one that realizes or satisfies it to the exclusion of any other incompossible physical object) but rather whether some physical object realizes or satisfies a unique function (that is, one that is realized or satisfied by a physical thing to the exclusion of any other incompossible function). A Lewisian version of our criterion would be something like: a physical thing *P* uniquely realizes a function or functional role *f* if, and only if, *f* makes the functional statement (i.e. the Ramsey sentence) about *P* realizing some *f* true and any substitution of *f* for any other function in that statement renders the statement false. Our argument will show that this biconditional is true because both sides are false. If *f* is a pure function and *P* a purely physical thing, then *f* cannot make the relevant Ramsey sentence true and neither can any other pure function which is substituted for *f*. No pure function can make the relevant Ramsey sentence true, and neither can any physical realizer if the functions are formal functions. This also means that, with respect to pure functions, such functions cannot be uniquely realized (both in our and Lewis's sense) by any physical thing whatsoever. For what it is for a pure function to be really realized is, in part, for it to be realized to the exclusion of any other incompossible pure function, and since (as we will show) no purely physical thing, change, or process can do that, no physical thing really realizes a definite pure form on account of its physical properties alone. So Lewis' notions of 'unique realization' and even plain 'realization' have no applicability to pure formal structures or functions.

4. KRIPKE AND PHYSICAL COMPUTING MECHANISMS

Now on to the thick and thin of the argument. Kripke attacks functionalism[130] at "its weakest rather than its strongest link" (Kripke, 1984, p. 1). This is the fundamental functionalist distinction between the abstract causal diagram or description of an abstract system, function, rule, or what have you, on the one hand, and the actual concrete physical system that is supposed to physically *realize* or *implement* that abstract diagram, function, rule, or description (cf. Kripke, 1984, p. 2). Alternatively, we can draw the distinction in terms of the abstract, pure function itself (e.g. the addition function, the NAND function) and the concrete physical mechanisms or components that purportedly realize it (e.g. neurons, transistors, circuits, etc.). There is, of course, no single way in which such a distinction is cashed out. It undergirds comparisons between "an engineer's hardware description of the actual physical setup of a computer and the abstract description on the highest level that an automata theorist would give" (Kripke 1984, p. 4). It likewise undergirds theories that are based upon the distinction between the higher-order, abstract description (e.g. a Ramsey-sentence), and the first-order physical processes that purportedly satisfy and fit that higher-order description.[131] The problem with functionalism is that, though it recognizes the distinction from the outset, it then proceeds to ignore it and obliterate it, ending up with an actually mistaken and incoherent account of both how we think mathematically and logically and how physical computing machines work.[132] (In effect, this view attributes to physical computing systems capacities and abilities it does not and cannot have, and attributes to us capacities and processes that we either don't have or a not constitutive of formal understanding. In short, we don't do what they do, and they don't do what we do.)

It is important to be clear on what exactly is the problem with the functionalist's distinction between the abstract diagram (or description) of an abstract functional system, state, or structure, which disparate physical systems can implement in different ways, and the physical systems themselves that are supposed to realize the abstract diagram or description (Kripke 1984, pp. 2-3). It is not that the distinction itself is problematic; indeed, it is not. It is a truism. Computability theory is a branch of mathematics, not of physics or computational engineering. The former concerns pure

---

[130] Recall that Kripke takes his argument against *computational functionalism* to be, if valid, applicable to any kind of view, including causal-theoretic functionalism. Our interest in Kripke's argument is that it establishes a far more theoretically profound and wide-raging claim than the refutation of a particular naturalistic theory: it establishes that no physical fact and regularity can be formally determinate. All of which is tantamount to saying that if a physical object realizes an abstract function, it cannot be on account of its physical features and properties alone. Yet we do that all the time when we think mathematically and logically. More on this in chapters 4 and 5.

[131] This is what Piccinini (Piccinini, 2017; cf. Piccinini 2015) calls "the problem of computational implementation" which, as he tells us, "may be formulated in a couple of different ways. Some people interpret the formalisms of computability theory as defining abstract *objects*. According to this interpretation, Turing machines, algorithms, and the like are abstract objects. But how can a concrete physical system implement an abstract object? Other people treat the formalisms of computability theory simply as abstract computational *descriptions*. But how can a concrete physical system satisfy an abstract computational description? Regardless of how the problem of computational implementation is formulated, solving it requires an account of concrete computation — an account of what it takes for a physical system to perform a given computation."

[132] "But I feel that this view [viz., psychological and intellectual states and activity as the physical realization of an abstract structure or function], considered as a philosophy of mind, embodies not only a mistake about persons but even in a way a mistake about actual computing machines or things that we call computing machines. It ignores the very distinction that functionalists themselves emphasize: that is the distinction between (1) the abstract diagram of an abstract mathematical automaton, and (2) the physical machines themselves that realize the diagram." (Kripke, 1984, p. 2) Notice that this is a particular formulation (concerning the adoption of automata theory in the philosophy of mind) of a more general and fundamental problem: that pure, abstract formal structures can be realized objectively and intrinsically by purely physical systems.

structures, the latter the physical implementation or realization of those structures. The problem lies in trying to explain our cogitative powers and acts (human computation) as nothing but the *physical* implementation or realization of pure (mathematical and/or logical) functions and rules. If, as our argument will attempt to establish, no purely physical thing can be said to realize intrinsically and determinately[133] a unique abstract function or diagram such as to exclude some other incompossible abstract function or diagram, then functionalism as a theory of the mind, and in particular as a theory of higher intellectual capacities and abilities, is bankrupt. For any φ (let φ be any physical system or process) said to realize some abstract function f, there will always be some other *incompossible* abstract function, g, which all the physical facts of φ equally well satisfy. From this and other considerations it follows that nothing intrinsic to φ (i.e., no interpretation-independent fact) determines which function φ realizes. Therefore, if functionalism is true, there's no interpretation-independent or observer-independent fact as to what functions or formal structures a given physical system realizes. Hence, functionalism, if true, entails the formal indeterminacy of the physical and, insofar as it is a theory of the mind and purports to explain formal understanding as the physical realizations of formal structures, it entails the formal indeterminacy of the human mind and therefore that there is no such thing as formal understanding. Note the implication: if functionalism (and, more fundamentally, type or token physicalism) is true, it implies that we do not have cogitative powers in any intrinsic and determinate way (for these are no more than the physical realizations of abstract structures, which cannot be physically realized in a determinate and intrinsic manner) but only in an extrinsic and observer-dependent way.

Though the considerations underpinning the argument are enough to refute functionalism, there is the risk of throwing the baby out with the bath water, moving on to bigger and better things, and neglecting to pursue the fundamental question of why functionalism fails on Kripke's watch. The reason as to why functionalist and computational accounts fail in the relevant respect is because they take it to be the case and intelligible to claim that human formal capacities, abilities, and activities are in some complex way the physical realization, implementation, or satisfaction of pure formal structures. Let me put this more clearly. The claim is that formal capacities (i.e., our capacity for formal understanding) are physical capacities, and that formal abilities are physical abilities and thus that formal abilities (i.e., abilities to engage in formal activity like calculating) are physical abilities for a certain kind of physical activity. Our task will be to show, on the basis of Kripke's argument, that this claim however construed (whether by means of the notions of physical realization, implementation, supervenience, or type-identity) cannot be true. Physical determinacy can't be formal and formal determinacy can't be physical. The rest will fall like a domino effect. The

---

133 The fact that no determinate abstract function is *intrinsically* realized by physical things and processes does not mean that no such function is *extrinsically* realized. I will come back to this at a later stage. Equivalently, from the fact that physical things are indeterminate with respect to pure functions or structures, it does not follow that they cannot *simulate* or *approximate* such functions or structures. I will also explain this in detail later on. For now, let me just say that the rejoinder "but any physical system will approximate or simulate incompossible functions, since for any φ that is taken to approximate or simulate ι, there will be some other incompossible pure function ι^* which φ will equally well approximates or simulates; hence, φ cannot simulate or approximate any pure function" will not do. For approximation and simulation are forms of *extrinsic realization*, and this implies that there cannot be such a thing as a case of approximation or simulation that isn't relative to an observer. The objection either confusedly and fallaciously attributes to us the claim that 'approximation' and 'simulation' are objective, non-relative, and intrinsic features of physical things––in which case the objection, if the claim was truly ours, would cut mustard—or it actually concedes our point by pointing to the fact that whether a physical object approximates or simulates a pure formal structure will be something *extrinsic* to the physical object itself, and relative to an observer, interpreter, designer, or what have you. This is what it means for any physical thing to realize a pure function: to approximate it or to simulate it, and these in turn are not intrinsic attributes of a physical thing. Formal approximations and simulations are not natural, i.e. real, features of the natural, physical world. They are *human (mind-)dependent features of things*.

reasoning underpinning our Kripkean argument is, naturally, far richer and more complex than what I've sketched thus far.[134] So let us go into the thick and thin of the argument.

Kripke's problem for functionalism can be stated in the following way: which function a physical computing system computes requires a bit of abstraction and is relative to its idealization; but which idealization is true of that physical system is, in turn, relative to the intentions of its interpreter (in the case of a designed computing mechanism like a computer, relative to the intentions of its designer(s)). In other words, neither the function a physical computing system purportedly computes nor the determinant of the correct and true idealization or interpretation of it are part of the objective and intrinsic physical causal structure of the system. Consider the following passage:

> …I may build a concrete machine, made of metal and gears (or transistors and wires), and declare that it embodies the function I intend by '+': the values it gives are the values of the function I intend. However, there are several problems with this. […] I cannot really insist that the values of the function are given by the machine. First, the machine is a finite object, accepting only finitely many numbers as input and yielding finitely many as output as output – others are simply too big. Indefinitely many programs extend the actual behavior of the machine. Usually, this is just ignored because the designer of the machine intended it to fulfill just one program… (Indeed, the appeal to the designer's program makes the physical machine superfluous; only the program is really relevant. The machine as physical object is of values only if the intended function can somehow be read off from the physical object alone.) Second, in practice it hardly is likely that I really intend to entrust the values of a function to the operation of a physical machine, even for that part of the function for which the machine can operate. Actual machines can *malfunction*: through melting wires or slipping gears they may give the wrong answer. How is it determined when a malfunction occurs? By reference to the program of the machine, as intended by its designer, not simply by reference to the machine itself… The last two criticisms of the use of the physical machine… – its finitude and the possibility of malfunction – obviously parallel two corresponding objections to the dispositional account.
>
> (Kripke, 1982, pp. 33-35)

---

[134] In footnote 24 (Kripke, 1982, pp. 35–37 n.24) , Kripke pointed out in a preliminary manner how the considerations that undergird the so-called skeptical paradox provide ample grounds for undermining functionalism. He then said that he hoped to elaborate on such remarks elsewhere (Kripke, 1982, p. 37). The argument to be discussed is the elaboration (Kripke, 1984, p. 1). Secondly, Kripke's argument against functionalism is an offspring of the anti-functionalist and anti-dispositional remarks he made in footnotes 24 and 25 of his seminal book on Wittgenstein (Kripke 1982, p. 35-40). It is not, however, a skeptical argument. Indeed, his unpublished anti-functionalist argument can be (and in fact *was*) elaborated and defended without commitment to the so-called "skeptical paradox", which occupied the bulk of that book. For his attack on functionalism has nothing to do with global or local skepticism about meaning, but with the fundamental presupposition of any functionalist theory, for instance: that a physical system, state, or process, can uniquely realize a 'abstract' function (i.e. to the exclusion of *any other* incompatible 'abstract' function). It is our claim that this undermines not only functionalism but any kind of physicalism about the realization of abstract functions because it shows that no physical thing on account of its physical properties alone can *intrinsically* and *determinately* realize a definite formal operation or function. Such functions are not only *not* physical functions (something with which the functionalist agrees), they are not even determinately realizable by them, except in an observer-dependent way (which, in the case of computational functionalism, would consist in the intentions of the machine's designer). This is what I take to be the crucial import of Kripke's argument. It extends further than functionalism and computationalism downright to any version of materialism about the cogitative or human computation.

Kripke here addresses two problems that play a crucial role in his anti-functionalist argument: *the finitude problem* and *the malfunction problem* (cf. Kripke 1984; Buechner, 2011). According to Kripke, any physical computing mechanism that can be said to compute or realize non-trivially a function will be

(1) subject to the laws and necessities of nature[135], and thus to decay, corruption, and wear;
(2) subject to malfunction[136], and
(3) subject to finite, temporal operative powers and life-time, both actually and potentially.[137]

---

135 'Physical laws' would have been misleading, as that label threatens to convey only the laws or necessities discovered and formulated by physics (including at the quantum level). But by the laws and necessities of nature we mean not only those but also the laws and necessities uncovered at the level of chemistry, biogenetics, neuroscience, and even psychology (if any). So I have deemed it better to talk about the laws and necessities of nature, rather than physical laws. I do take psychological laws to constitute a species of laws of nature (though not reducible to other "lower-level" species of laws of nature like the laws of physics) except in the case of rational psychology. For the psyche of the rational animals can be subject to *formal laws*, which cannot in any intelligible way be a species of the laws of nature.

136 Physical mechanisms are subject to malfunction, but they are not capable of making *formal* mistakes or erring *formally*. What about animals? Are they also subject to error? What is an "error"? Are they synonymous with mistakes or with malfunctions? There is a conceptual difference between being subject to eventualities in the manner of (1) and eventualities in the manner (2). Such eventualities as suffering corruption, wear, and decay are not so much cases of malfunction and error as necessities of nature—say, states which sooner or later any object of nature succumbs to on account of its materiality. Of course, wear and decay may give rise to malfunction, but they are not the same thing. Cases of malfunction and error in nature (e.g. chromosomal abnormalities) are deviations from definite "norms" of nature that cause the material object (or some subcomponent of it) not to develop, act, or behave in accord with its natural function, thus frustrating to whatever extent its proper functioning, the actualization of its natural potentialities, and in some cases its natural fulfilment and even existence. Of course, cases of (2) can bring about, or accelerate the coming about of, (1) in a given material substance or object, and vice-versa (e.g. old age osteoporosis). Now, the notions of "malfunction", "error", and "deviation", are spoken of in many ways, and the way which we are interested in the most for present purposes is that which is spoken of in relation to pure formal norms or rules, rather than in relation to "natural functions" or "norms of nature" (including psychological "norms" or "laws" if there are any, i.e. psychologism). A way to introduce the gist of Kripke's argument but in line with these considerations would be to ask whether natural processes, changes, and states can count as "formally" deviant (i.e. deviations from pure formal rules or pure functions, and not just from a natural forms or laws) in any *intrinsic*, non-observer dependent, way. (As we will see in the next two sections, "error", "mistake", "malfunction", and "deviation", are norm-dependent features and concepts, that is they are dependent on a norm. So a "norm", be it natural, psychological, formal, social, etc., is epistemically, conceptually, and ontologically prior to a correct or incorrect, i.e. deviant, instance of it.) It is hard to see how this can be answered in the affirmative without a reduction of formal features to natural features, and thus of pure forms to natural forms. Functionalism is precisely an attempt to answer in the affirmative without a reduction. Supervenience physicalism is likewise another attempt. Our argument refutes them both.

137 This is important and bears a stark resemblance to Fodor's considerations in favor of the *productivity* of thought (Fodor, 2008, 1998, 1983, 1975; Fodor and Pylyshyn, 1988). Human (formal) understanding seems unlimited and unbounded in ways no material thing is, at least in principle. Of course, in actuality humans only ever do finitely many things, they only judge and reason finitely many times in finitely many ways, the only generate finitely many syntactic structures, etc. But this is a limitation due to a host of material factors like finite lifespan, finite memory capacity, being awake, tiredness, fatigue, etc., that have nothing do to with the formal operation as such but with its materially bounded execution. It is not by a "law of thought" that human understanding produces finitely many instances of syntactic, logical, and mathematical types of operations. It is a "law" of nature, of matter, to which we are subject on account of being rational, intellective *animals*. But our cogitative powers as such do not seem to possess any such limitations, which is why human understanding can be *in principle* carry out, say, infinitely many recursive operations (as Chomsky would formulate it). Even if not potentially infinite in its forms or ways of understanding, such that the resources for structuring and guiding thinking were finite, what humans could generate and grasp on the basis of that finite set of resources would still be potentially infinite. Aside from the constraints placed on humans by materiality, there seems to be no upper bound to what humans can understand and do, formally speaking (I ignore issues concerning Gödel's disjunction, particularly those arising out of the assertion that the human mind is an effectively axiomatized formal system). In sum, material constraints are placed on the potential infinity of cogitative acts in which human beings can *in principle* engage in. Interestingly, neither Fodor nor Chomsky see this as undermining appeals to our species-specific biological endowment as sufficiently determining our capacity for generating a potential infinity of discreet (cogitative)

None of these are true of (indeed, none are intelligibly ascribable to) pure formal rules and functions. But aren't they true of and ascribable to their applications or computations? Yes.[138] As animals we are equally subject to (1), (2), and (3). But if whatever satisfies (1), (2), and (3) cannot realize a pure function, then neither can we. We will leave this conundrum for chapters 3 and 4. For now, we simply consider the fact that if anything physical really applies or computes, say, an infinite arithmetical function, it will do so only in a finite amount of time, through finitely many steps, and *in the world*, being thus subject to the necessities and laws of nature, and being capable of realizing the function until it either decays, malfunctions, or reaches the limits of its finite computational capacities. This is true of us as material, embodied beings. So here's the dilemma: either we do not apply definite formal rules and therefore cannot have formal understanding (e.g. we do not calculate, compute, reason validly or invalidly, etc.) or we are capable of acting and thinking in ways that are formally determinate and thus cannot do so *solely* on account of our nature as embodied beings (put less strangely if possible, we can transcend the limitations of matter so as to 'achieve' formal determination). In chapter 4, I will argue that the human capacity for formal understanding does transcend such limitations though the exercise of that understanding *qua* ability exercised (activity) is constrained and subject to such necessities insofar as any human activity is necessarily embodied or bodily in some sense (say, dependent on our neurobiological endowment as a necessary condition).[139] But it's only its underlying material and psychological mechanisms, which are a

---

acts. But Chomsky comes close at times to our position. "…human language: in particular, that "marvelous invention" of a means to express our thoughts in novel and limitless ways that are constrained by our bodily state *but not determined by it*; that are appropriate to situations but *not caused by them*, a crucial distinction; and that evoke in others thoughts that they could have expressed in similar ways – a collection of properties that we may call "the creative use of language." (Chomsky, 2002, pp. 49–50; my emphasis)

[138] In our adaptation of Kripke's argument, we do not claim that our applications (which are concrete and can take time and effort to carry out, can be slow or fast, etc.) of formal rules have no material cause. It is we, rational animals, who do mathematics and logic and therefore are spatio-temporally, physiologically, and psychologically constrained in our formal activities. If the relevant cogitative powers were in no need of a material basis there would be no explaining why we *need* some material forms of representation, like suitable notation, and why some are more appropriate than others (e.g., Arabic numerals vs. Roman numerals) for acquiring and furthering formal understanding, for the development of pure formal sciences, and for the study and progress of non-pure (e.g., physical) sciences. It would leave unexplained why we need a material *means* of representation (e.g., signs taken as symbols, sensory phenomena, etc.) in order to think or understand formal concepts at all. But the means of representation cannot be identified and confused with that which the means of representation are said to represent on pain of a category mistake (that is, on pain of conflating a representation with what it is a representation of). What we are saying, however, is that the material component does not cause nor determine as a *sufficient* condition the formal determinacy of the act. At a purely material level of analysis, the form is lost. The material aspects of the activity considered purely materially satisfy incompossible formal rules. But this just means that human formal activity is not a purely material activity, and therefore neither can the capacity and abilities for that sort of activity be purely material. Once one realizes this, the problem of teaching formal rules and communicating formal understanding dissipates, for one is no longer under the illusion that actions are nothing more than their material basis and means of representation. How exactly this is supposed to work, metaphysically, is something beyond the scope of the present work. We highlight this issue, however, in the conclusion.

[139] I have said that (2) is not true of human formal understanding *as such*, but only true of material things. Yet humans do err and make mistakes when engaged in numerous activities, and formal activity is not the exception. Moreover, some formal mistakes might be due to some neurological malfunction. To what then are we to attribute formal mistakes, misunderstandings, confusion, and errors? The suggestion behind this objection is that it would seem that it would have to be attributed to their matter, and that we would have to appeal to something like the limitations that matter imposes on formal operations in order to make our case intelligible. Since our argument implies that no material thing, process, or feature can be formally deviant *in se*, and thus that there cannot be any such thing as formal errors or formal mistakes in/by the physical (e.g. the brain cannot reason validly, but it cannot reason invalidly either; it simply is not the sort of subject to which formal validity can be attributed to), it follows (so the objection goes) that humans make mistakes because of their materiality or corporeality. But then it would seem that material things *are* suitable subjects of formal mistakes since any deviation from a formal norm or function in the case of human thinking would be attributed (or

necessary basis and condition for it, that satisfy (1), (2), and (3). Perhaps formal understanding and activity is some sort of hybrid between the physical and the formal, similar in that respect to the computational powers of physical computing mechanisms (see chapter 3, §7) but with the very important difference that the hybridity of human computation is intrinsic rather than extrinsic (for more on the claim the formal aspects of physical computation are *extrinsic* to it, see chapter 3). We leave that for another chapter, as it is a separate issue. Here I just want to press home the point that (1), (2), and (3) provide positive grounds, as opposed to mere intuitions about the physical, for making an argument in favor of the claim that no physical thing can *determinately* realize a unique pure function.[140]

Consider now the addition and quaddition functions. Take any physical mechanism, organic or artificial, at whichever level of matter you like, and ask 'what about it could possibly determine whether the addition or quaddition function is being realized in any given case?' [141] In general, the

---

attributable) to their matter. I am inclined to say that is all confused. Material processes and mechanisms can no more make formal mistakes (e.g. reasoning invalidly, misapplying modus ponens) on account of their materiality than it can make correct application of a given formal rule (e.g. reasoning validly, applying modus ponens). When I say that (2) is not true of formal understanding as such I mean that there is no such thing as 'understanding malfunctioning' (that makes no sense), but there is such a thing as a malfunction (which is not the same as a mistake or a confusion) of the nervous system (e.g. neuro malfunctioning) that can impinge both on the acquisition, retention, and exercises of relevant formal understanding. But this is not necessary in order for us to be confused or ignorant about a formal concept, or for us to be mistaken about and for having misunderstood a given formal rule. So formal mistakes, errors, misunderstandings, and confusion cannot be equated with the mechanical or physical notion of malfunction. In addition, it is *we* not some part of us that gets conceptually or formally confused (e.g. compare the use of 'confusion' in 'conceptual confusion' and 'muscle confusion'). It is I, not my brain or my nervous system, who gets an 18 (an A+) on a mathematics exam and who is judged on its formal competency. Yet without a brain and nervous system, no such competence could take place. I am not denying that embodiment is a necessary condition for human formal understanding; I am denying its sufficiency.

[140] A common misunderstanding of Kripke's argument (as I understand it) consists in thinking that the argument entails the absurd claim that the physical is indeterminate *tout court*. That is a misunderstanding of numbing proportions. The Kripkean argument does not imply that the physical is indeterminate, only that it is *formally* indeterminate. It does not gainsay that physical structures and processes like benzene rings and photosynthesis are determinate, only that they are *formally* so. Physical structures are *not* abstract, pure mathematical or logical structures. There is a categorial difference here. What is true is that we do use and construct formal systems to *model* and *study* physical systems, their properties, and processes. But such formal systems and features are *idealizations* of the physical systems, their properties and processes, and cannot be intelligibly identified with them. Plus, a good ideal of convenience and human interests play a role in the selection or construction or preference of some idealizations over others (this need not entail anti-realism; this just points to the fact that science is a human enterprise and activity, and therefore inevitably subject to all of the features of the human condition). On a separate note, one ought not to confuse issues of formal indeterminacy with a function's being undefined for certain values. A mathematical function that is undefined for certain arguments *is* a determinate function, which is why we can talk about it, define it, calculate according to it, and come to know that it is true of it that it is undefined for some *n*. But a formally indeterminate function makes *no sense* at all, since what it is to be a function or a formal rule is, in part, either to be or to determine an invariable component of a certain sort that is dispositive for any possible transformation of the right kind of variable component. To say that modus ponens is formally indeterminate is to say, well, nothing, for it amounts to the claim that modus ponens is not modus ponens. A formally indeterminate function is no pure function at all: *it makes no sense*. But a function that is undefined for certain arguments is a perfectly intelligible function and quite commonplace in mathematics.

[141] According to functionalists, there are *in principle* indefinitely many different physical ways in which logical and mathematical functions could be realized. On the causal-theoretic functionalist view of realization (see, e.g., Block 1980; Melnyk 2003, 2006), the realizer of a function in a given physical system would be whatever physical (i.e. neural) property or process plays the causal role (say, is responsible for producing the correct outputs on the basis of arbitrary inputs and other states) associated with the relevant abstract function. But Kripke's problem implies precisely that the same set of neural or physical processes can be said to satisfied incompossible abstract functions. Sydney Shoemaker's (Shoemaker, 2007) definition of the notion of property realization, which he calls the 'subset view of realization', is equally subject to Kripke's problem and cannot be used to dissolve it. For "as a first approximation, property P has property Q as a realizer just in case (1) the forward-looking causal features of property P are a subset of the forward-

burden consists in finding a physical property or process that can realize a unique pure function *f*, such that for any *f** it realizes, *f** = *f*. The physical fact would have to be one that does not violate (DR), since violating (DR) leads to a contradiction (we will show this in the next chapter when we come to 'the conjunctive' and 'the disjunctive' objections to our argument). Kripke's argument can almost be formulated as a test: take any physical thing whatsoever, constrain its interpretation however much you like so that it cannot be said to compute or realize a function trivially *à la* Putnam, select the function it implements, then run Kripke's argument and cash in on the indeterminacy.

At the beginning of the previous section, we quoted a passage about a computer that is supposed to compute the addition function but cannot exclude incompossible functions from being realized. There we were confronted with the problem that the physical computing machine can be interpreted equally well as either an (malfunctioning) *adding* machine, a (properly functioning) *quadding* machine, a (properly functioning) *skadding* machine, and so on indefinitely. What's interesting about this scenario is that even when considering computations over '57', issues about formal indeterminacy abound. For whether the numerical output '5' that the computer gives for any arguments greater than '57' is a mistake or not, a malfunction or not, a wrong answer or not, *depends on which rule or function the computer is (supposed to be) following*. Thus, the output cannot be used to settle the matter, for it cannot constitute a determinant of the function being carried out. Furthermore, since we can always let '57' represent a number too great to be computable in a finite number of steps, or too great to be processable by anything finite, we would see that nothing in the actual physical and behavioral properties of the system can determine intrinsically a unique function. The Stabler-Lande variation of the argument will bring home this point in a much more incisive way.

Before proceeding, however, I feel it necessary to make my position extremely clear on the following issue. I am not arguing that there are no such things as roles or functions in the physical realm. I am arguing that whatever these roles or functions are, whatever their nature, they cannot be *abstract* and *formal* (and thus, their being satisfied or realized by some physical being cannot be a matter of the physical realization of abstract phenomena, nor of the abstract phenomena supervening on the physical phenomena). For when we abstract away from the given material properties of the phenomenon at hand, what we are left with is something categorially different from that with which we started and which cannot be identified with the original phenomenon, much like the abstraction of a point and a line, of a circle and its radii, from a written dot and wooden wheel, for example, gives rise to categorially different phenomena, which cannot be said to be identical with nor determinately and perfectly realized by the written dot and the wooden wheel even if the geometrical concepts of a point and line were abstracted from them: the dot and the wheel only realize the geometrical concepts imperfectly, and in an observer-dependent way.

5. PHYSICAL COMPUTING MECHANISMS AND THE INEVITABILITY OF FORMAL INDETERMINACY

looking causal features of property Q, and (2) the backward-looking causal features of P have as a subset the backward-looking features of Q" (Shoemaker, 2007, p. 12). But any subset of the causal features of Q that can be said to consist of the causal features of P, can likewise be said to be said to consist of the causal features of some other incompossible function P*.

Let us now formulate Kripke's argument by means of Edward Stabler's own counter-argument to Kripke's anti-computationalist lectures (Stabler, 1987).[142] We introduce the concept of a finite state automaton (FSA), which is just one of several mathematical models of computation. But our discussion should give an idea of how it generalizes to all other kinds of models of mathematical and logical computation. To Stabler's formulation of Kripke's argument we add Lande's variant of it (Lande, 2011). The Stabler-Lande formulation brings quite clearly and elegantly to the fore what Buechner calls "the deep problem for functionalism", which is that "the abstract diagrams of an abstract mathematical automaton are only imperfectly realized in physical computing machines (PCM). One cannot read off the function computed by the PCM from its physical causal behavior." (Buechner, 2011, pp. 345-346)

Consider a very simple electrical circuit that is supposed to compute a determinate pure function, namely the infinite identity function.[143] The circuit diagram looks like this:

input _____ output

This is an extremely simple physical device. It consists only of a wire that conducts voltage pulses from an input voltage point to an output voltage point. All we need is "a wire connecting the input point to the output point" such that when a voltage pulse (of such-and-such duration and magnitude) is applied at the input point on the conductor (with such-and-such properties), a voltage pulse (of the same duration and magnitude) will be transmitted to the output point of the conductor. Furthermore, that the conductor if engineered correctly and functioning properly will do so is *guaranteed* by the laws of physics.[144] Interpreting each such pulse as one digit in our unary notation, Stabler goes on to state that "the circuit can be seen as realizing the infinite identity function on the natural numbers." (Stabler, 1987, p. 9)[145]

Let us stipulate, with Lande (Lande, 2011, p. 3), that the actual conductor of Stabler's physical computing device is such that it decays after a sequence of two voltage pulses is administered. Once a sequence of two voltage pulses is administered, the conductor "ceases to transmit a voltage to the output point". Call the physical device Γ. (That Γ's wire decays and ceases conduct after a sequence of two voltage pulses is administered is just a case in point—much like '57' for the quaddition function—since one can always make the number of sequences of voltage pulses

---

[142] Computer scientist Edward Stabler was present in one of the conferences where Kripke gave his (as-of-today) unpublished refutation of functionalism. Stabler's paper (Stabler, 1987) is entirely based on the talk Kripke gave at that conference, and it was the only article to have been published on Kripke's anti-functionalist argument until (Buechner, 2011). Stabler's article is discussed in (Lande, 2011) too.

[143] I have chosen the Stabler-Lande exposition of the argument over Buchner's because Lande takes into account the *internal* structure of the physical object at a given point in time, whereas Buechner considers explicitly only the function and the behavior of the physical system. Thus, Lande avoids more readily the accusation that his is an argument against behaviorism, rather than functionalism (though, it must be said, that some have argued that functionalism inevitably collapses into logical behaviorism; see Putnam, 1988). It does not take much to see how Kripke's plus/quus considerations and Buechner's informal exposition both apply to functionalism. Whatever the case, the following version of Kripke's argument is ultimately a hybrid, incorporating elements from Kripke's own formulation of the argument (Kripke, 1984), from Buechner's (Buechner, 2011), Stabler's (Stabler, 1987), and Lande's (Lande, 2011).

[144] "To eliminate all unnecessary complexity, let's consider a device which is much simpler than an adder, but like it in that it realizes an infinite function on the natural numbers. […] All we need is a wire connecting the input to the output. The required background condition is just the presence of the conductor between these two points, because we can presume that the laws of physics tell us that if a voltage pulse (of such-and-such duration and magnitude) is applied to the input point on the conductor (with such-and-such properties) a voltage pulse (of such-and-such duration and magnitude) will be transmitted to the output point after some finite amount of time." (Stabler 1987, pp. 8-9)

[145] "However, our conductor will last (i.e. continue to have the requisite properties) for only so long, so it will not be able to perform the computation on very large numbers." (Stabler 1987, p. 9)

processable by the Γ before it decays greater, and without any change in the implications of the argument, and because any physical or material substance, device, change, or process will cease to be or to properly function at some point. For matters of simplicity, we will let that point be '2'. Second, we do not need to commit ourselves to the actual conductor decaying after a certain sequence of two voltage pulses is administered; we could just stipulate that the conductor is such that it cannot process anything greater than two voltage pulses, and thus never emits any voltage for any sequence of voltage pulses greater than two. This is likewise in line with the nature of physical or material things, since there will always be logical and mathematical problems too great and too complex for anything physical to carry out). We will, however, stick with Lande's example, leaving it as an exercise for the reader to rehearse (if she so wishes) the argument in the way just suggested.) Now, according to Stabler, Γ computes the addition function. But according to Lande, let us say, Γ computes another function, called the two-counter (or quidentity) function.

Consider the following two finite state automata[146] (FSA, for short):

(1) an identity-computing FSA which, for each sequence of n inputs, produces a sequence of n outputs;

(2) a two-counter-computing FSA which, for a sequence of n inputs, where $n \leq 2$, produces a sequence of n outputs, but when $n > 2$, produces a "null" output.[147]

---

[146] A finite state machine or finite state automaton (FSA; plural: finite state automata) is, ontologically speaking, a pure formal structure; specifically, it is a mathematical model of computation. There are other mechanical models of computation such as combinatorial-state automata (CSAs), but what we have to say does not hinge on the differences between the two. Our argument applies equally well to CSAs. Now, an FSA is specified by means of the specification of a set of inputs, a set of outputs, a set of internal states and a set of state-transition rules which determine the possible transformations of input to output on the basis of the internal states of the FSA and the relations that obtain amongst those internal states. For example, on the basis of some arbitrary specification, one can construct an FSA which given input I_2 and internal state S_0, will produce output O_8 and transition to internal state S_1, all in accordance with specified rules regarding the transformation of input states to output states on the basis of the internal state of the FSA at a given time and the internal state transitions that ought to take place for the relevant transformations to happen. One can now see, roughly, how an FSA could be said to carry out or compute a given mathematical or logical function. (There are limits of course to FSAs. But this is not to the point. Remember: our target is the claim that there is such a thing as the physical determinate realization of a given formal, e.g. mathematical or logical, structure. FSAs are mathematical structures; so any claim to the effect that physical systems can implement or realize FSAs determinately will be subject to our critique, regardless of what the computational theory says about the computational powers of FSAs.) The view that human beings are FSAs was borne out of the early functionalist papers of Hilary Putnam, who thought human beings were physical realizations of Turin machine tables:

> A 'machine table' *describes* a machine if the machine has internal states corresponding to the columns of the table, and if it 'obeys' the instruction in the table in the following sense: when it is scanning a square on which a symbol s_1 appears and it is in, say, state B, that it carries out the 'instruction' in the appropriate row and column of the table (in this case, column B and row s_1). Any machine that is described by a machine table of the sort just exemplified is a Turing machine.
>
> (Putnam, 1960, 1975a, p. 365; cf. also Putnam 1967/1975a, pp.433–4)

But that makes no sense, for finite state automata and Turing machines are abstract, mathematical objects. What he should have said is that we are physical implementations of finite state automata (Kripke, 1984, p. 5). But this just grants the categorial differences that obtain between a mathematical object and a physical object. Putnam later abandoned and staunchly criticized this theory as a theory of human psychology, but that the physical implementations of an FSA could realize determinately and uniquely that FSA was put into question by Putnam for reasons entirely different from ours.

[147] It is not clear whether Lande takes the two-counter function, in itself, to be a unary function that is undefined for any $n > 2$ or rather a unary function that is defined for any $n > 2$, such that for any $n > 2$, $n \bullet 0$. A physical analog device without the capabilities for numerical representation would not be able to discriminate between the two. No transmission of voltage can be compatible with both. But if we allow for digital, numerical representation of the number of voltage pulses at both the input and at the output points, the "null" output (which is, by the way, a symbolic output)

.
.
.

(The three dots are meant to convey that the list of functions or FSAs attributable to Γ is indefinitely long. But, for our purposes, (1) and (2) will suffice.) The problem should be obvious by now: which FSA is true of Γ? Which FSA, and thus which function, does Γ realize? If Γ computes the identity function, i.e. if Γ is an identity-computing FSA, then its ceasing to transmit any output after $n > 2$ voltage pulses will manifest a malfunction, a break down, or a limitation. But if Γ computes the two-counter function, i.e. if Γ is a two-counter-computing FSA, then its not transmitting anything after $n > 2$ voltage pulses will *not* manifest a malfunction, nor a limitation of any kind, but rather *correct* behavior, i.e. it will manifest that it is a properly functioning two-counter computing device.[148]

The problem is straightforward: for any sequence of $n \leq 2$ voltage pulses, it seems that none of the physical properties of Γ determine whether it is an identity computing device or a two-counter computing device. And whether Γ's behavior for any sequence of $n > 2$ voltage pulses manifests a malfunction/limitation or properly functioning behavior depends on which function Γ is supposed to be computing. Yet nothing intrinsic to Γ determines which![149] But perhaps we can do

---

could be used to represent the first option, while the "0" output for any $n > 2$ the second option. Of course, even with these constraints, determinacy problems still abound, for whether the "null" or the "0" output is the *correct* value of the relevant inputs rather than a mistake or a case of malfunction and limitations will, again, depend on the function the physical device is supposed to compute. So, again, we cannot on the basis of the actual outputs of the physical system go on to determine which function it computes because the actual outputs of the system, which are finite, can only count as a subset of the set of possible (potential and counterfactual) values of that function, and such a subset will be equivalent to a subset of the set of possible values of an incompossible function.

[148] Again, we are taking into account only these two functions. If we include in our elaboration of the argument other incompossible functions (say, a function that is undefined for n if $2 < n < 57^{28}$, such that for any n, if $n \leq 2$ or $n \geq 57^{28}$, n; undefined, otherwise) then whether Γ is functioning correctly, i.e. behaving in accord with the rule or function it is supposed to follow or compute, when its wire ceases to conduct will depend on exactly which one of the functions it is supposed to be computing. Also notice that we can build something like Γ for computing some logical functions. For instance, we can interpret Γ as computing a tautologous equivalence relation, letting the wire represent the biconditional '*iff*' and letting the natural numbers, onto which we map the voltage pulses of Γ, represent the propositional letters P, Q, R, S… such that the set of all propositional letters {P, Q, R, S, P´, Q´, R´, S´, P´´,….} is at the least countable. Thus, one voltage pulse means P, two voltage pulses means Q, and so on, and such that for any propositional letter ξ, ξ. We can now proceed to define an incompossible function, say, the ququivalence relation, according to which for any propositional letter $\xi \leq 2$, ξ; undefined otherwise. The same problem arises: Γ can be said to satisfy both. Likewise, with a little bit of engineering wit and skill, one can refashion Γ so that its physical processes and behavior can be interpreted as a radian-to-degree measure computing FSA, according to which for any radian measure as input (e.g. π), the equivalent value in degree measure is given as output (e.g. 180^{o}). We can then go on to define a number of incompossible functions or FSAs that Γ will equally well satisfy. There's no limit to the number of incompossible functions one can point to at each level of complexity of the physical system. Of course, in order to become non-trivially interpretable as a physical system that computes more complex logical (and mathematical) statements, the physical system will have to be more complex and mechanically rich than Γ, but the point will still remain since it is not a matter of simplicity or complexity but a manner of *category*, and thus of the categorial differences between the formal and the physical.

[149] Stabler recognizes the problem (his paper is, after all, an attempt at refuting Kripke's argument against computational functionalism). After having drawn the aforementioned circuit diagram, Stabler goes on to say: "However, our conductor will last (i.e. continue to have the requisite physical properties) for only so long, so it will not be able to perform the computation on very large numbers. […] The physical facts of the example are clear enough: for a time the wire will transmit pulses, but then it will fail to do so. Suppose we start applying pulses and the wire ceases to conduct after the 57^{th} pulse. We wanted to describe this wire as the realization of an infinite identity function, but it seems that we could just as well say that it was a realization of the function which maps numbers less than 58 into themselves and thereafter maps every number into 5 – on this story, like the first one, the function was computed correctly up until the

better that this "intuitive" sketch of Γ's situation. Let us focus, not merely on Γ's behavior at some particular time *t*, but also on Γ's physical structure and configuration at *t*. We will again follow Lande in formulating a "state description" of Γ at *t* which is a description of all the relevant properties of Γ together with all the causal laws that apply to Γ at *t*.[150] We will see that even given a maximally specific physical description of Γ, no physical aspect of Γ can determine which FSA it realizes, and thus which function it computes.

Before going into Lande's exposition, let us state first Stabler's (Stabler, 1987, p. 11) refined definition of what it is for a physical computing system to realize an abstract structure or function. We say that a physical system computes a function *f* if, and only if,

(1) there is an "interpretation" or "realization" function *In* which maps a set of finite sequences of physical 'input' states of the system onto the domain of *f*, and an interpretation function *Out* which maps a set of finite sequences of physical 'output' states onto the range of *f*, such that
(2) physical laws guarantee that (in certain circumstances *C*, and *if the system satisfied conditions of normal operation N for long enough*), if the system went successively through the states of an input sequence *i*, it would go successively through the states of the corresponding output sequence *f* where Out(*f*) = *f*(In(*i*)).

We will examine this definition in detail later, but note that if we want the interpretation function to be applied with precision we need to ascertain a maximally specific description of the physical system in question in order to map the input states of the physical system to the domain of a given function and the output states of the system to the range of that function. We need to specify which states of the physical system count as its input states and which as its output states, and which count as its computationally or functionally relevant states (e.g. those states that process the input and generate the corresponding output). As Stabler rightly observes, "Saul Kripke's suggestion [is] that even given appropriate constraints on interpretation, indeed, even fixing precisely the interpretation of physical states, there is an additional problem with computational accounts. In many cases, even when the interpretation is fixed, it is still indeterminate what function a physical system is computing. The problem arises most clearly for infinite functions (that is, for functions with infinite domains), but it is not restricted to these." (Stabler, 1987, p. 2)

Suppose we wish to express this in terms of the realization of an FSA. Let us use lower-case letters for labeling the states of the physical system and upper-case letters for the abstract states of the FSA. We assign the labels $s_1, s_2, \ldots, s_n$ to the states of the physical system that are related to each other in certain specified ways (the so-called "internal states" of the system), and the labels $i_1, i_2, \ldots, i_n$

---

device broke and ceased to satisfy the conditions of normal operation. Indeed, why not say that the wire was a realization of every function that agrees with the identity function on actually computed values, up to 57? In every such case, the wire was computing the function correctly up until 57th pulse, at which point it "broke down". But this view conflicts with *any* claim to the effect that any system computes one function rather than another (finite or infinite) function which agrees with the first only (finitely many) actually performed computations." (Stabler, 1987, p. 9)

[150] "In the first place, we want to base our judgment on the structure of the system at a particular time *t*, and not simply on its behavior. […] In any case, relying only on the behavior of the system to determine what FSA it realizes risks collapse into behaviorism. The suggestion, then, will be that we look at Waldo's structure at a particular time. This will consist of taking the description of all the properties of Waldo at t (a so-called "state description" at t), in conjunction with all the causal laws that apply to Waldo. On the basis of this maximally specific physical description of Waldo, and the relevant causal laws, we should be able to come up with counterfactual conditionals specifying how Waldo would behave under various conditions. From this, we can build up a machine table for Waldo, on the basis of which we can determine which FSA it realizes." (Lande, 2011, p. 3) Lande calls his physical computing device 'Waldo'.

and $o_1, o_2, \ldots, o_n$ to the inputs and outputs respectively. We can then go on to say rather imprecisely that Γ realizes an FSA F if and only if there is a one-to-one correspondence or mapping between the states, the inputs, and the outputs of Γ and the states $S_1, S_2, \ldots, S_n$, the inputs $I_1, I_2, \ldots I_n$ and the outputs $O_1, O_2, \ldots O_n$ of F, respectively.[151]

Let us follow Lande (Lande 2011, pp. 3-4) again in considering two incompossible interpretations of Γ.[152] Take the first interpretation, which we will simply call "interpretation A":

(a) Let i_1 = an electrical pulse administered to Γ.
(b) Let o_1 = an electrical pulse output by Γ.
(c) Let the internal states of Γ be s_1 = no electrical pulse is being conducted by Γ, and s_2 = an electrical pulse is being conducted in Γ.

As Lande rightly points out, given this interpretation, we are now in a position to map the states of Γ onto the machine table of an identity-computing FSA. According to interpretation A, when the first voltage pulse, i_1, is administered, Γ goes into state s_2 and produces output o_1, after which it then goes back to s_1 (we let s_1 be the initial state of Γ). When two voltage pulses i_2 are adminitstered are administered, Γ goes into state s_2 and produces output o_2, after which it then goes back to, and so on. That Γ does not actually transmit any voltage after two voltage pulses is just a fact about the material limitations of Γ as a concrete device, "a feature of how the FSA is implemented in with this particular shoddy condutcor." (Lande, 2011, p. 3) But if the wire in Γ were not to decay (or, in Stabler's jargon, if Γ were to satisfy conditions of normal operation for long enough), Γ would satisfy interpretation Id, and thus continue to "exhibit all the proper state-transitions of an identity computing FSA", and could therefore be said to actually compute finitely many values of the infinite identity function.

Consider now a different interpretation of Γ, call it "B". Let "Stage 1" be Γ's initial condition; "Stage 2" is that stage in which Γ is after the first voltage pulse is administered; "Stage 3" that stage which Γ is in and remains in indefinitely after the second pulse. We can now go on to give interpretation B (Lande, 2011, p. 4):

---

[151] Chalmers characterizes concrete (i.e. physical) computation in terms of the notion of implementation, which he defines thus: "A physical system implements [realizes] an FSA (with input and output) in a given time-period if there is a mapping f from physical states of the system onto formal states of the FSA and formal inputs and outputs to the system onto inputs and outputs of the FSA such that: for every state-transition $(I,S) \rightarrow (S',O)$ in the specification of the FSA, if the physical system is in a state s and receiving input I such that $f(s) = S$ and $f(i) = I$, this causes it to transit into a state s' such that $f(s') = S'$ and to produce output o such that $f(o) = O$." (Chalmers, 1996, pp. 321) Chalmer's definition of what it is for a physical system to compute a function places the necessary constraints on the interpretation function (i.e., "a mapping f") of the physical system so as to avoid problems of triviality, as raised by Searle and Putnam. But Chalmer's definition cannot avoid Kripke's problem. Chalmer's account of concrete computation or implementation is a species of the isomorphic and causal-counterfactual account of physical computation, and therefore equally subject to the argument developed in the main text. Note that these formulations of concrete computation are not strictly speaking instances of what Godfrey-Smith (Godfrey-Smith, 2009) called the "simple mapping account" of computation (but see Piccinini, 2015, pp. 16–25), which we will come back to when discussing the semantic account of computation (chapter 3, §4).

[152] Though we've left it, for the most part, at an intuitive level, it must be noted that each interpretation should be understood as coming with a built-in realization or implementation function that allows us to systematically map the input states, internal states, and output states of the physical system Γ to the input states, internal states, and output states of the abstract FSA which Γ is supposed to realize. The abstract machine state diagram constitutes an effective procedure for calculating the values of the relevant function. This mapping is, at bottom, equivalent to Stabler's mapping of the states of Γ to the domain and range of the function.

(a) Let i_1 = an electrical pulse administered to Γ.

(b) Let o_1 = an electrical pulse output by Γ, and o_2 = no electrical pulse output by Γ.

(c) Let the internal states of Γ be

1. s_1 = no electrical pulse is being conducted in Γ, and Γ is in "Stage 1";
2. s_2 = an electrical pulse is being conducted in Γ, and Γ is in "Stage 1";
3. s_3 = no electrical pulse is being conducted in Γ, and Γ is in "Stage 2";
4. s_4 = an electrical pulse is being conducted in Γ and Γ is in "Stage 2";
5. s_5 = no electrical pulse is being conducted in Γ, and Γ is in "Stage 3";
6. s_6 = an electrical pulse is being conducted in Γ, and Γ is in "Stage 3."

(d) After the transition from $s_1 - s_4$ (i.e. after the first two pulses/stages), Γ will only ever switch between s_5 and s_6.

Thus, given interpretation B, the states of Γ can now be mapped onto the machine table of a two-counter or quidentity-computing FSA. That Γ ceases to conduct after the second voltage pulse is wholly in accord with what interpretation B would represent a concrete implementation of a two-counter-computing FSA as doing.

The crucial question at this point is whether the physical configuration and causal properties of Γ alone can exclude any one of the two incompossible interpretations from being equally well satisfied and thus determine which one is true of Γ. The answer is 'no'. Neither interpretation violates the total physical structure and causal makeup of Γ. If, say, Γ's ceasing to conduct after two voltage pulses counts as an accidental feature of a concrete physical implementation of an FSA, then Γ will be in accord with A[153], but if it's a feature of the functional or computational configuration of an FSA, then Γ will be in accord with B. But that is precisely what is in question! For what makes a feature of Γ accidental or not (a mistake, a case of malfunction, of deviant behavior, or not) is dependent upon which function or FSA Γ is supposed to be realizing. For what counts as accidental and deviant is consequent upon what counts as essential and correct. Thus, if the physical structure and causal properties of Γ cannot determine which function or FSA it computes, it cannot determine which features are accidental or essential and thus, cannot determine what counts as a malfunction or what counts as normal conditions of operation. We will have time to come back to this when dealing with Stabler's objection. For now, we point out that which physical stages of Γ are accidental or part of its functional organization is determined by the function or FSA Γ realizes, not the other way around.[154]

---

[153] But so with indefinitely other incompossible interpretations, in which case neither accord nor conflict would occur here—as Wittgenstein would have said. For even if we grant that the wire's decay or its not conducting anything after two voltage pulses is an accidental feature of the physical implementation of some FSA *F*, say an identity-computing FSA, there will be at least one other incompossible FSA *G* which will likewise be compatible with the wire's decay being an accidental feature of its concrete physical implementation. The same goes for taking the wire's decay or its not conducting anything as part of the functional or computational configuration.

[154] Lande expresses a similar point as follows (I've changed the wording of Lande's passage to make it coincide with our text): "Will a maximally specific physical description of Γ at t tell us whether Interpretation Id or Interpretation Quid is the right interpretation of Γ —whether one "cuts deeper" into Γ's causal topology than the other—and therefore whether Γ is an identity-computer or a two- counter? No. Both interpretations respect the physical structure of Γ. Moreover, both interpretations respect causal laws about Γ: (1) with respect to Interpretation Id, it is true that if Γ were to continue indefinitely, as if it were in Interpretation Quid's "Stage 1," it would continue to compute identity; (2) with respect to Interpretation Quid, it is true that if Γ were to continue indefinitely, the sequence of inputs would cause, in a law-like fashion, Γ to transition through stages until it continued to produce a "null" output. The only question is

A natural response would be to say "simply check the design!", i.e. look at the designers' instructions for Γ and you'll be able to determine what Γ was intended to compute. If it was built as an identity-computing conductor, then we'll know that the decay of Γ's conductor is a *defect* or *limitation*, an accidental feature, of a concrete physical implementation of the identity-computing FSA, and therefore Γ computes the identity function. But if it was built as a two-counter computing conductor, the decay of Γ's conductor would actually serve to realize or implement in a particular way the computational or functional configuration of a two-counter FSA, and thus Γ would have been computing the two-counter function all along. This response is correct, but is one that concedes Kripke's point: that the physical structure and causal makeup of the system *alone* are not the determinant of, and thus cannot suffice to determine, which function the system realizes. Two designers, say an Id designer and a Quid designer, can build the same physical device, say Γ, to compute the identity function and the two-counter function respectively. Γ would satisfy both equally well. If Γ fell from the sky, with no instruction manual whatsoever, there'd be no telling which function the system computes. Designer Id and designer Quid would be equally well justified (or unjustified) in attributing to the fallen Γ their respective incompossible functions. As Kripke himself writes,

> "If a machine fell from the sky as a physical object, there is no saying anything about what it means [and does]. Two different designers might design identically the same physical object with opposite intentions. The same object may validly represent the distinct intentions of two designers. […] [Even] without any funny labeling of states and without anything strange or unusual, it is also true that a machine never literally realizes its state diagram. That is something that involves a certain amount of idealization and abstraction. Idealization and abstraction are perfectly all right, but they cannot be said to be part of an objective causal theory of the machine in the sense that another theory which conflicts with them will be objectively false."
>
> (Kripke, 1984, p. 9)

To say that an instruction manual or the abstract design for Γ is what ultimately determines which function the concrete physical device Γ realizes just grants that Γ, *in se*, cannot satisfy (DR), which is precisely what we wanted to prove.[155] One needs to look to something other than Γ (e.g. instruction manual, intentions of the designer) to settle which function Γ (or any other physical system for that matter) realizes. And even then, two incompossible designs or instruction manuals can be said to apply to Γ.

One can try to fix the situation by amending interpretation A, for example, so as to allow "the causal description of the machine to count its wearing out and breaking down as so to speak part of what was intended." (Kripke, 1984, p. 8) Suppose we amend it by assuming that the physical system, Γ, goes on forever without any malfunction. This is an idealization of the physical device. We can then proceed to specify what Γ would do if it were in such counterfactual, idealized

---

whether the stages of Γ are an accidental feature of its concrete implementation, or a feature of its functional organization. In other words, what features of Γ count as part of the relevant causal structure, which is to be mapped onto an FSA, and what features are accidental to the implementation? Two different counterfactual stories could be told, on perfectly legitimate grounds, given all the properties of Γ, and all the relevant causal laws, depending on which interpretation we latch onto. And there appears to be no reason to prefer one story to the other. This is a real problem when we realize that the "stories" are supposed to be metaphysically accurate descriptions of Γ." (Lande, 2011, p. 4)

[155] Lande thinks that this only proves that Γ cannot satisfy the weaker (UR), not the stronger (DR). Dillard would seem to reply much the same way. We will come back to this in the next chapter, when we consider objections to our argument.

conditions. But this won't solve the problem. For this amounts to nothing more than to conceiving Γ without ever deviating from the function it is (or was) computing. But which function (if any) Γ is (or was) computing is what is in question! In addition, this move is a performative concession to our point that whether something counts as *deviant* depends, ultimately, on the interpretation (and normative requirements thereof) that is agreed upon or chosen for the physical system. "Deviations," writes Kripke, "from this [or that] behavior are deviations from a normative requirement which is being placed on the machine" (Kripke 1984, p. 8) and which is not part of the actual, objective, physical causal configuration of the physical system. In other words, whether something in a computational system[156] counts as a malfunction (that is, a deviation from a norm) is dependent upon which function or FSA the system is supposed to realize, and this in turn is dependent upon a (designer-)given interpretation. In short: one cannot solve the problem by pointing to a malfunction, then abstracting from it and idealizing it away, because one cannot know whether something is a malfunction unless one already knows what is the *norm* (i.e. the function or FSA) to which the physical system is supposed to adhere. Of course, if we know which function or FSA Γ computes, we'll know how to idealize it. But the whole point is that there is no such observer-independent fact to be known! Since both abstraction and idealization depend on us taking for granted which function the system computes, they cannot be used to settle which function Γ computes. Therefore, idealization and abstraction of Γ (and any other physical device for that matter) cannot be used to refute Kripke's argument, for they presuppose what is in question. But suppose one responds, instead, by idealizing Γ such that the wire never decays (without any commitment to whether this is a malfunction or not). What would it be doing? We will come back to this when considering Stabler's counterfactual objection to Kripke's argument.

This is our problem: if Γ satisfies interpretation A, then it is true that if Γ were to continue functioning indefinitely without malfunction (as if it were on "Stage 1" of interpretation B), it would compute arbitrary values of the identity function. If Γ satisfies interpretation B, then it is true that if Γ were to continue functioning indefinitely without malfunction, it would compute arbitrary values of the two-counter function by transitioning through stages until arriving at one which continuously produces the "null" output. The wire's ceasing to transmit voltage will be either a case of malfunction or a case of proper functioning depending on which description and function Γ realizes. But the physical structure and causal properties of Γ alone cannot exclude either incompossible function and description, and thus cannot determine which function (if any) is being realized. To paraphrase Quine, these are logically incompatible but physically equivalent.

A number of objections can be raised at this point. To these we now turn.

---

[156] Note that it makes no sense to talk about an abstract, i.e. mathematical, computational system malfunctioning or functioning properly. There's no such thing as a mathematical function malfunctioning (or functioning properly). Such things are just not the sort of features attributable to mathematical structures, and to do so would be to incur in a category mistake. Events such as malfunctioning, breaking down or wearing out can only be said of the physical systems that are supposed to realize such structures. This is a further example of a categorial difference between physical structures and formal structures.

3
The Kripkean Challenge: Objections and Replies

In this chapter we consider some of the main objections that can be raised against Kripke's anti-functionalist argument.

1. THE CONJUNCTIVE AND DISJUNCTIVE OBJECTIONS[157]

Consider the following objection:

(1) **The conjunctive objection**. Why can't we just say that Γ satisfies both A and B, and their respective interpretation function? We could argue for the following conjunction: Γ is an identity computing device and Γ is a two-counting device. What follows from the physical and causal properties of Γ is that both functions are being computed by Γ.

Let us call this view *conjunctvisim* about formal operations. The motivation for this response seems to be that if a physical system cannot exclude two incompossible functions from being realized at the same time and in the same respect, then it should be taken to realize and compute both. Why can't Γ conjointly realize both FSAs and thus the concrete computation of both functions? Our answer is pretty straightforward: because such as thing is unintelligible and therefore cannot constitute a real possibility. To claim otherwise leads, logically, to absurdity.

Suppose Γ conjointly realizes the identity function and the two counter-function.[158] Then Γ insofar as it realizes a definite pure function is determinate with respect to which rule-counterfactuals are true of it. For instance, if Γ computed the identity function, then the following counterfactual would be true of Γ:

01) If Γ had received a sequence of three voltage pulses at its input terminal, and if Γ had not malfunctioned, Γ would've produced a sequence of three voltage pulse at its output terminal.

But if Γ computes the two-counter function, then the following counterfactual would be true of it:

02) If Γ had received a sequence of three voltage pulses its input terminal, and if Γ had not malfunctioned, Γ would've *not* produced a sequence of three voltage pulses (because would've produced no sequence of voltage pulses at all) at its output terminal.

Generally, if Γ is an identity computing FSA, then if Γ were to receive a sequence of n voltage pulses, where $n > 2$, and if Γ is not malfunctioning, Γ would emit a sequence of n voltage pulses at its output terminal, but if Γ is a two-counter FSA, then if Γ were to receive a sequence of n voltage pulses, where $n > 2$, and if Γ is not malfunctioning, Γ would *not* emit a sequence of n voltage pulses at its output terminal.[159] The consequents of the two counterfactuals are contradictories. They

---

[157] See Lande for his take on what I call the conjunctive objection and the disjunctive objection (Lande, 2011, pp. 5-7). I have encountered the disjunctive objection most often when discussing Kripke's argument with other philosophers and colleagues, but I have seldom encountered the conjunctive one. I consider it nevertheless.

[158] Note that what we say applies to *any* formal rule and function.

[159] There is a problem here. The counterfactuals seem to be *descriptive* (about what the system would do or would've done) rather than *prescriptive* (about what the system ought to do or ought to/should have done). But this is misleading.

The antecedent of such counterfactuals requires that the physical system does not deviate from acting in accord with the function it is supposed to realize (i.e. that it does not malfunction and make a mistake when computing). Thus, if Γ is supposed to compute some function *f* and if Γ does not deviate from acting in accord with the function *f*, then Γ will compute arbitrary values of *f*, and thus Γ would compute what it ought to or should compute. But the ways in which Γ can be said to compute what it should or ought to compute are not univocal. In terms of the function *f*, if Γ is supposed to compute some function *f* and if Γ does not deviate from acting in accord with the rule for computing *f*, then if Γ were given the appropriate kind of input it would compute arbitrary values of *f* and its computations would be *correct*. But to say that Γ is under some rational, moral, social, or pragmatic obligation to compute correctly arbitrary values of the *f* function does not make sense. *We*, of course, may be under such obligations (e.g. 'to pursue truth'), and we can opt not to obey them, but pure function or formal rules though norms are not themselves obligations, imperatives, or commandments. In the case of logic, formal rules constitute the truth-preserving ways of reasoning. In the case of mathematics, formal rules constitute the correct ways of calculating, say, what the ratio of an arc to a radius in a circle is. In the case of grammar, formal rules (e.g. morpho-syntactic rules) constitute, in part, what is or can be a well-formed grammatical expression. Regulative rules concern such pragmatic and prudential imperatives as whether we ought to think in truth-preserving ways, whether we ought to follow rules for grammatically well-formed sentences, and so forth––that is, rules that regulate activities that are the activities they are independent of those rules. For example, if we build into Γ the capacity for taking into account practical or pragmatic considerations, which will at times override Γ's purpose of computing arbitrary values of the function *f*, one can then go on to say that though Γ produces in certain definite contexts answers that are not correct (e.g. not in accord with the *f*-function) it's answers in those contexts are the ones it ought to produce (say, because, given its program, pragmatic or prudential concerns demand it). With respect to physical devices, therefore, an anti-normativist view (Boghossian, 2005, 1989; Davidson, 1992; Gluer and Wikforss, 2009; Millikan, 1990; Wikforss, 2018) is somewhat on the right track when it is qualified in the following way: physical systems are subject to formal rules in a regulative rather than a constitutive manner (the distinction between regulative rules and constitutive rules can be found in (Searle, 1970), and it is often appealed to by anti-normativists (Kusch, 2006)). But even this should not be understood univocally with how we are subject to regulative rules. I mean that physical computing systems are subject to *designer expectations and constraints*, and whatever application the notion of "norm" or "rule" has on these kinds of systems will therefore be observer-dependent, imposed on them from without. But with respect to human reasoning, however, pure formal rules and functions are constitutive, rather than merely regulative. Kusch *et al* claim that only regulative rules are normative. But this is a misunderstanding of 'normative'. Constitutive rules are standards of correctness and as such determine what is or counts as a correct (and incorrect) application of the rule. In this sense, constitutive rules are normative. Formal rules do determine what counts as an application (and, therefore, as a misapplication) of the rule, but they do not necessarily *compel* anyone to apply them. They are not commands. That the formal rule of modus ponens determines whether a particular inference is a case of it and therefore formally valid, has to do with what modus ponens *is* and nothing with whether it is convenient or wise or praiseworthy to draw that inference in a particular context. Likewise, what counts as a touchdown in American football is dependent, and constituted, by the rules of the game which are standards of correctness. Thus, if a player runs out of bounds, and the referee calls it a "touchdown" the referee either misspoke or does not understand the rules of the game (e.g., the rules for what counts as a touchdown). The referee would be wrong in her call or ruling but not because she is breaching some sort of social etiquette or pragmatic consideration (actually, all the same, she might be acting in accord with regulative exigencies of social etiquette, if one imagines a situation in which calling the player who went out of bounds at mid field as having scored a touchdown is deemed a respectable and honorable act, or manifesting pragmatic wisdom, if one imagines a situation in which if she does not rule it a touchdown, her family's life will be in jeopardy) but because she is misapplying (or, worst case scenario, misunderstanding) the rule for touchdowns. Similarly, that inferring "Miami is in the US" from "If Florida is in the US, then Miami is in the US" and "Florida is in the US" is a case of modus ponens and thus a formally correct or valid inference *is not* a matter of pragmatic or prudential commandments. But aren't logical rules both constitutive and regulative? We do not need to settle this here. They are surely constitutive, but whether they can also act a regulative in a non-arbitrary way depends in turn on the whether human beings, qua *rational* animals, have a telos in which the pursuit of truth figures. But this is neither here nor there. The preceding considerations may lend support to the claim that logical rules have to be constitutive rules. It is modus ponens, *qua* formal rule, that determines what counts as an instance of modus ponens. The rule constitutes a way of valid reasoning, and thus what it is to reason in accord with it. What I *can* do is opt out from complying with the regulative rules that surround such intellectual practices in which the application of modus ponens (or any other logical rule) is necessary or deemed essential and a rebel against such practices and decide not to reason in accord with deductively valid forms of reasoning. But this says absolutely nothing about what follows from a conditional and its antecedent, nor about modus ponens as a standard of correctness. The old adagio 'regulative rules can be breached, constitutive rules cannot' carries some weight if understood correctly. Constitutive rules can be abandoned (see e.g. non-classical or non-standard logics) to some extent, but they cannot be

cannot both be true at the same time and in the same respect. The supposition that Γ is a case of both an identity computing FSA and a two-counter computing FSA leads to a *reductio ad absurdum*. Therefore, Γ cannot conjointly realize both functions.[160] The conjunctive objection fails.

One can reformulate the conjunctive objection so as to claim,

> 03) Γ realizes an identity-computer and it realizes a two-counter, but in any case in which Γ is a properly functioning two-counter, it is a malfunctioning identity-computer, and vice versa. (Lande, 2011, p. 5, n. 9)

Moderate conjunctivism, as we shall call the sort of conjunctivism expressed in 03), grants that both functions cannot be realized at the same time and in the same respect. So the way the moderate conjunctivist maneuvers her way out of the incoherence of conjunctivism is by saying that when Γ stops computing one function, it computes the other incompossible one.[161] In fact, unless this is some sort of *ad hoc* response, the fundamental claim is what when a physical computing mechanism stops computing some function, it begins to compute some other (incompossible) function. But this is obviously flawed. First, the whole point of the Kripkean argument is that there is no physical process, change, or property of Γ that determines in which case Γ is a properly functioning identity-computer and in which case a properly functioning two-counter computer. Consequently, there is no physical objective fact of the matter that determines in which case Γ is a malfunctioning identity-computer and in which case a malfunctioning two-counter computer. The moderate conjunctivist's

---

breached though they can be misapplied (e.g., a case of affirming the consequent is not a case of breaching or violating modus ponens; it's a case of not following a valid form of reasoning and following an invalid one instead). Constitutive rules are not imperatives, though one can juxtapose imperatives to it (e.g., "humans ought to reason in accord with deductively valid forms of reasoning").

[160] Again, we are arguing that two incompossible pure functions cannot be realized by anything *at the same time and in the same respect*. This qualification must be kept in mind at all times when considering our argument, otherwise the argument might get confused for an argument against the possibility of parallel processing and simultaneous computations. We are not arguing against quantum computation. That is why Γ is an excellent example to make our point, because it is a simple conductor endowed with the properties for one computational task only. There are no subcomponents and parts in Γ computing different functions and realizing different programs. If it helps, think of Γ as an atomic part or subcomponent of a larger system and whose purpose is to compute a single definite function. Then run Kripke's argument. Nothing will change. Relatedly, the Kripkean argument concerns not merely physical wholes, but any physical part of a physical whole that may be appealed to as a possible realizer or determinant of a definite and unique pure function. This should be clear. To object "all Kripke's argument shows is that no physical whole can determine a unique pure function, but not that none of its parts can" is to misunderstand the argument. All one need do is carry out Kripke's argument at each subcomponent level of the physical with the same the result. We can take the brain as a whole, or we can take parts of the brain, as a potential determinant of a definite formal operation or structure. We can even go down to the level of individual neurons and cells. The claim is that no matter at which physical level we formulate the argument, the physical will satisfy incompossible pure formal structures and operations. (Note my claim in chapter 1 that the argument is about the *kind* of facts the functionalist appeals to, i.e. *physical* facts, not about some subset of that kind.) Functionalism or computationalism requires that the human mind (at least with respect to doing logic and mathematics) be the physical realization of abstract formal rules and functions. But since nothing physical can realize pure functions, functionalism *qua* realization physicalism is not tenable. Whether we consider the physical realizer, say Γ, either as a whole system or as merely a part or component of a whole system does not matter. If it's physical, it will be formally indeterminate. This can be proven by letting the objector choose any part of the physical whole she likes. Once she selects the physical part whose physical configuration is supposed to realize a definite formal rule *R*, we run Kripke's argument on that physical part with an incompossible formal rule *R**. The part will satisfy incompossible pure functions, and formal indeterminacy will abound. If the part is not "atomic", and is composed of still further parts, we can repeat the dialectical process. At every stage, even at the stage of atom decay, Kripke's argument will apply.

[161] So Γ is always computing some function? How can the moderate conjunctivist make sense of the fact that a physical computing mechanism may *stop* computing if it can always be taken to compute some other function whenever it stops computing the originally intended one?

response presupposes that it *is* or *can be* physically determinate when it is a properly functioning (or malfunctioning) identity-computing system and when it is a properly functioning (or malfunctioning) two-counter computing system. It begs the question. A conjunction is true only when both conjuncts are true. But we show that neither conjunct can, on account of the physical properties of Γ alone, be true. Not only is each conjunct of a function incompossible with the function of the other, each function of each conjunct has in turn some other incompossible function which anything physical that satisfies one satisfied the other. Moderate conjunctivism presupposes that we can settle definitively by appeal to the physical properties of the system alone when Γ is a case of a properly functioning identity-computing system (and thus a malfunctioning two-counter computer) and when Γ is a case of a properly functioning two-counter computing system (and thus a malfunctioning identity-computing system). But this is precisely what is in question. Nothing in Γ can make that moderate conjunction true.

Second, the list of incompossible functions or rules satisfied by the physical system at any given time is not limited to just *two* incompossible functions, but rather is indefinitely long, and it will lead to extremely convoluted conjunctive formulations about the functions a physical system realizes, such as

04) Γ realizes a function *f* and it realizes a function *g* and it realizes a function *h* and it realizes a function *j* and it realizes a function *k*…

where none of the functions is identical with any other. Let us suppose that the functions *f, g, h, j,* and *k* are incompossible functions which are all satisfied (like addition, quaddition, and skaddition in our original example) by the physical, causal configuration of Γ. For any input less than two voltage pulses, Γ can be said to be (to conform to the rules for) computing *f,* and *g,* and *h,* and *j,* and *k*. Therefore, Γ would count as a properly functioning *f-er, g-er, h-er, j-er,* and *k-er* at the same time and at the same respect. After two voltage pulses, Γ would count on the basis of the very same physical conditions at the same time and in the same respect as a malfunctioning *f-er,* a malfunctioning *g-er,* a malfunctioning *h-er,* a malfunctioning *j-er,* and a malfunctioning *k-er*. We know this yields contradictory results (see above). But the important thing to note is that the grounds or conditions that determine whether Γ is a properly functioning or a malfunctioning *f-er* (which are partly determined by the function it computes, that is, *f* in this case) *cannot* be the same that determine whether Γ is a properly functioning or a malfunctioning *g-er*—and so forth for *h, j,* and *k*—if there is to be any intelligibility to the claim that physical properties alone can determine whether a definite function or FSA is being realized and realized properly (or improperly).[162] Otherwise it is

[162] This is similar to the point made by (Buechner, 2011, p. 355-356) as to why assigning malfunctions and normal conditions of operations each to equivalence classes will not work: "There is no finite bound on the distinct kinds of errors that any PCM [physical computing mechanism] might make because there are indefinitely many distinct physical processes that can result in making an error. *Assigning equivalence classes to all errors that output number n categorizes the differences in causes of errors as the same*. Outputting *n* because a microship overheats and outputting *n* because of a conformational change in a molecule are not the same thing, even though each outputs *n*. One designer might take the hot microship to be a malfunction, but the conformational change in the molecule to be normal. For each physical processe a PCM undergoes that is an error relative to the intentions of one designer, there is another designer for whom it is not an error, but a normal condition of operation. Thus the error-inducing physical processes a PCM undergoes cannot be assigned to an equivalenc class, since what is an error is not an intrinsic condition of it, but is relative to the intentions of its designers." (my emphasis) We agree. What's more, if any error that the mechanism can be said to incur in can be construed as part of the normal conditions of operation under some other incompossible interpretation which the mechanism satisfes, then assigning equivalence classes to all errors that output number *n* not only categorizes the differences in causes of errors as the same but it categorizes the differences in causes of normal conditions of operation as the same, since *n* as an error under some interpretation counts as correct under some other incompossible

indeterminate which function is computed. Yet the physical conditions are *ex hypothesi* held constant. Hence, the physical conditions cannot suffice to determine a definite conjunction of incompossible functions since the very same physical conditions cannot discriminate between incompossibles (they cannot even discriminate between incompossible conjunctions, say between 'conjunction' and 'qunjunction').

The bottom line is this: all the physical processes and properties of Γ whenever it is functioning 'in accord' with some pure function ψ (e.g. '…but in any case in which Γ is a properly functioning ψ-er…') are *also* compatible (1) with Γ being 'in accord' with the normal (or abnormal) conditions of operation of an indefinite number of incompossible functions and (2) with Γ being 'in conflict' with the normal (or abnormal) conditions of operation of an indefinite number of some other incompossible functions. For example, let *f* be the addition function. Suppose Γ is re-engineered so as to count as a physical mechanism which computes the addition function. Then, suppose Γ begins to produce '5' for any computation over pairs of inputs greater than '57'. An 'addition designer' would say that Γ is either malfunctioning and therefore no longer computing the addition function or is a mechanism that computes an entirely different function. But a 'guddition designer' would say the exact same thing with respect to her computational design: Γ is either malfunctioning and therefore no longer computing the guddition function (where guddition is a function that, for any pair of number less that 57^{10} yields the very same values as the addition function but for any pair of numbers greater than 57^{10} randomly outputs a prime number) or is either a mechanism that computes an entirely different function. Let *f** be the guddition function. According to these designers, if Γ *is* malfunctioning then it would be at the same time and in the same respect both a malfunctioning addition mechanism and a malfunctioning guddition mechanism. But this is a contradiction. The exact same considerations can be seen to apply to the 'quaddition designer' and the implementation of the quaddition function by Γ. Let *g** be the kuaddition function (a function which (i) for any pair of numbers less than 57 gives you the very same values as quaddition, and (ii) for any pair of numbers greater than 57 but less than 125^{5757} yields the very same values as quaddition, i.e. 5, but which (iii) for any numbers greater than 125^{5757} yields 0). Therefore, for any pair of numbers less than 57, whenever Γ is said to be properly functioning *f*-er it could be said to be a properly functioning *f**-er, a properly functioning *g*-er and a properly functioning *g**-er. Suppose it does not explode or melt or whatever, and it goes on to compute numbers greater than 57. Then in that case when Γ is said to be an improperly functioning *f*-er it could also be said to be an improperly functioning *f**-er, but it could also be said to be a properly functioning *g*-er and a properly functioning *g**-er. And so on indefinitely at any stage.[163] It does not

---

interpretation. This is absurd. Suppose, further, that outputting *n* instead of *m* due to some physical processes an conditions *U´* (e.g. an overheating of a microchip) is said to be an error according to some FSA for the implementation of some rule or function *f*. *U´* will, relative to some other FSA for implementation of an incompossible rule or function *g*, count as part of the system's normal conditions of operation. But *U´* will also count as an error according to some other function *f** (see next paragraph in main text, and n. 163) incompossible with *f*, and as part of the normal conditions of operation according to some other function *g** incompossible with *g*.

[163] Suppose further that the functions *f´, g´, h´, j´,* and *k´* are functions incompossible with *f, g, h, j,* and *k*, respectively, but in such a way that whenever Γ is a properly functioning *f*-er, Γ is also an improperly functioning or malfunctioning *f´*-er, and when Γ is properly functioning *g*-er, it is also a malfunctioning *g´*-er, etc., and conversely. Let *f* be the addition function. Let *f´* be a function which for any pair of numbers less than 57 gives you their square root, but for any number greater than 57 gives you the very same values as addition. Then, for any pair of numbers less than 57, Γ can be said to compute *f* and to miscompute *f´*. This is generalizable: every computation of some function is a miscomputation of some other function, and vice versa. Γ will at the same time and in the same respect an improperly functioning *f´*-er whenever it is a properly functioning *f*-er, and conversely. So if Γ goes on to compute numbers greater than '57' and yields what would seem to be their sum, it could be said to have gone from an improperly functioning *f´*-er to a properly functioning *f´*-er, and from a properly functioning *f*-er to an improperly functioning *f*-er. (And let us not forget that *f´* itself has an

matter where we draw the specific point of differentiation between incompossible functions. Indefinite number of incompossibles will be satisfied at every concrete moment of alleged divergence or conformity.

There is no limit to the number of incompossible functions that the system will satisfy, whether we construct its behavior at a given moment in time as one of proper functioning or as one of malfunctioning, and this would be so even if Γ were to continue without break down or wear forever.[164] Thirdly, the conjunctive objection makes it impossible to account for the possibility of *real* malfunctioning in a physical computing mechanism (e.g. that the physical computing mechanism ceases to compute anything at all on account of a design error) and cannot accommodate, particularly in the case human formal activity, the conceptual truth that failing to a apply a given rule is not necessarily succeeding at applying another rule, and vice versa. That is to say, moderate conjunctivism cannot account for the phenomenon of miscomputation; actually, it makes it impossible! Moderate conjunctivism could be made true, however, if the truth-conditions and truth-makers of the conjunction are dependent on the intentions of the designer, or on the interpretation imposed by some competent observer. Thus, whatever counts as a malfunction (and as an incorrect application or the absence of an application of the relevant rule) on one interpretation, can also count as a proper function (and thus as a correct application of a rule) on another interpretation. But this would relative to an Id-cum-Quid designer or observer. The conjunction would not be true of Γ in virtue of Γ's properties alone, which will be compatible with the intentions and interpretation of a Id*-cum-Quid* designer or observer (that is, a designer with an incompossible conjunction to the one proposed by the moderate conjunctivist for Γ).[165]

---

indefinite number of incompossible functions that yield the same values as it for any numbers less than 57, but not for all after 57, only for some. So, again, no limit to the functions Γ can be said to be miscomputing). When we put together these additional considerations with those of the main text, the incoherence should be obvious: whenever Γ can be said to be a properly functioning φ-er, it can be said to be a Γ is a properly functioning φ*-er, but also an improperly functioning φ'-er and so on. It makes contradictory statements true of Γ. We will see this more clearly below.

[164] See our reply to the Stabler-Chlamers account of computation. We explain there, following Ross, why not even an *actual* infinity of computations can determine which function the physical system intrinsically realizes.

[165] The reformulation cannot escape the inconsistencies of the original conjunctive objection (see Lande, 2011, pp. 5-7). This paragraph also points to a problem that every account of physical computation seem incapable of answering: the problem of miscomputation. This will become clearer when we get to Stabler's reply to Kripke's argument. But the idea is that causal, semantic, dispositionalist, and counterfactual (e.g. the Stabler-Chalmers approach) accounts of physical computation all fail to account for miscomputations. For whether a given computation of a physical computing system that diverges from the function it was programmed to compute is not something that is physically decidable, since it is not something that the material properties of the system determine. We would need access to the *program*. As we made clear, a computation of a given physical system that at some time *t* diverges from a function *f* for some given input *i* is compatible at *t* with its being a divergent computation of some other incompossible function *g*, and of some other incompossible function *h*, and of some other incompossible function *j*, and so on indefinitely. Likewise, the system's computation at *t* can also be said to be a (non-divergent, i.e. correct) computation of some other incompossible function *f'*, and of some other incompossible function *g'*, and of some other incompossible function *j'*, and of some other incompossible function *h'*, and so on indefinitely. In simple English: any physical case that 'satisfies' the conditions for being a miscomputation relative to some function satisfies the conditions for being a computation relative to some other incompossible function—by which I mean, relative to an indefinite number of incompossible functions; for there is no definite function the physical "computation" can be a computation of and no definite function the physical "miscomputation" can be a miscomputation of—; similarly, any physical case that satisfies the conditions for being a computation of some function satisfies the conditions for being a computation of some other incompossible function(s). Therefore, even if one says that a divergent computation is really the computation of a different function (causal accounts of computation are forced to say this, it seems, in which case there cannot be any such thing as miscomputation under a causal account; a huge problem) it would not be physically determinate which function that is. Piccinini (Piccinini, 2017) gets close to the crux of the problem in his defense of mechanistic account of physical computing systems: "Another feature of the mechanistic account is that it accounts for the possibility of miscomputation — a possibility difficult to make sense of under other accounts. To illustrate the point, consider an

The conjunctive objection proposes a violation of (DR). Recall our statement of the principle:

> *Determinate Realization (DR)* = For any realizer *P*, *P* realizes a pure function *f* if and only if the features of *P* relevant for realizing *f* determine that it realizes *f* and only *f*, such that for every *f** that *P* realizes, *f** = *f*.

It is precisely because the conjunctive objection violates (DR) that it leads to absurdity. For pure functions are determinate not only with respect to their actual cases, but also with respect to their potential and counterfactual cases. A genuine application of the addition function in arithmetic excludes its being equally a case of some other incompossible function (e.g. quaddition) not merely in its actual application, but also with respect to what any potential and counterfactual application. No physical thing, we argued, can be that definite. Pure forms are normatively dispositive for any possible or admissible transformations of their variable component, and whatever *really* realizes them must partake in that determinacy. To suppose that two incompossible pure functions can be realized at the same time and in the same respect is to commit oneself to a contradiction. It cannot be coherently countenanced. Thus, (DR) cannot be violated by a genuine realizer of a formal structure and should not be violated by any theorist wishing to explain the real application of formal rules and our understanding of formal concepts, functions, and rules. But the conjunctive objection stipulates that Γ violates it. Hence, if Γ violates (DR), Γ cannot determinately realize *in se* any pure function. Plus, bear in mind that even if we suppose *per impossible* that Γ satisfies a conjunction of incompossible pure descriptions, we can always construct another conjunction incompossible with the former that the system satisfies, and so on ad infinitum. Conjunction is a logical operator or form and thus a pure formal structure, in our technical sense. It would remain that none of the physical properties and processes of Γ can make a conjunction, or determine which conjunction is, true of it—except relative to the intentions of the designer or some observer-dependent fact.

Suppose the proponent of conjunctivism were to switch strategies and propose *disjunctivism* instead. Consider the following objection:

(2) **The disjunctive objection**. Why can't Γ just be said to satisfy the disjunction: "Γ computes the identity function or Γ computes the two-counter function"? All that follows from the physical and causal properties of Γ is that either the identity function is being computed by Γ or the two-counter function is being computed by Γ.

Prima facie, this looks like a more plausible objection to Kripke's argument. This objection claims that the same physical properties or system, say Γ, realizes a disjunction or a 'disjunctive property', call it β (e.g., '$f \vee g$') namely that of being either an identity computing system or a two-counter computing system. It realizes either function at any given time, but not both. Thus, Γ is such that

---

ordinary computer programmed to compute function *f* on input *i*. Suppose that the computer malfunctions and produces an output different from *f*(*i*). According to the causal (semantic) account, the computer just underwent a causal process (a manipulation of representations), which may be given a computational description and hence counts as computing some function *g*(*i*), where *g*≠*f*. By contrast, according to the mechanistic account, the computer simply failed to compute, or at least it failed to complete its computation correctly. Given the importance of avoiding miscomputations in the design and use of computers, the ability of the mechanistic account to make sense of miscomputation may be an advantage over rival accounts." A mechanistic account of computation—which is to my opinion the only viable account of physical computation—cannot avoid our conclusion either. The only thing missing in a mechanistic account of computation is a recognition of extrinsic denomination: physical computational systems are formally determinate only relative to something extrinsic to them, namely agreed upon interpretation.

there are some possible worlds in which it computes the identity function, and other possible worlds in which it computes the two-counter function. But this objection is confused. First, this objection doesn't account for what could make the disjunction true at any given moment. The physical properties of Γ, for example, will not be able to determine at any given time one disjunct over the other. For the physical properties are compatible with either (say, whether as case of a proper computation or as a case of miscomputation), so it is not clear what could make that disjunction true about Γ (or any other physical system). It seems that no truth-values could be assigned to the disjuncts. Secondly, as with the conjunctive objection, each disjunct can be parsed into a further disjunction consisting of two incompossible functions and whose disjuncts can in turn be further parsed into another disjunction of incompossible functions, and so on without end. For example, if one says that Γ realizes either *f* or *g*, then whenever it is said to realize *f* it could also be said to be realizing either *f* or qu-*f*, or either *g* or qu-*g*, and so on for each disjunct over and over again. Indeterminacy problems arise again. The potential physical determinants of the truth of the disjunctive property will always leave its truth-value undetermined and thus undecidable. Thirdly, let us not forget that 'disjunction' is a formal operation, and therefore one can always develop an incompossible disjunction, call it qu-β (e.g., '*f* Ⓥ *g*'), that Γ will equally well satisfy. Fourthly, recall that the number of incompossible functions satisfied by Γ is indefinitely great, and thus the disjunction would have to be indefinitely long, perhaps infinitely so. It is not clear then how such a disjunction would have any determinate value unless one presupposes that one of the disjuncts holds determinately in this or that world on account of some physical property of Γ. But this is a text-book example of a *petitio*.

Fifthly, as Lande correctly points out, anyone who espouses (2) has to renounce supervenience. For (2) implies that the same physical system could be either a case of an identity-computing FSA or a case of a two-counter computing FSA. But supervenience entails that no two things that are physically identical can be functionally (or mentally) different, for the functional properties supervene on (are uniquely and necessarily determined by) the physical properties. According to supervenience, there can be no change in the functional or formal aspects of a physical system without change in the physical properties and aspects of the system (cf. Kim, 1993; see e.g., Leuenberger, 2008). But the objection rests on the claim that the very same physical system can be either computing the identity function or computing the two-counter function, and thus can realize different functions in different possible worlds *without change in its physical properties*. But this contradicts supervenience. Furthermore, as we have already noted, in any possible world in which Γ can be taken to compute the two-counter function there will be at least one other incompossible function that the physical-causal structure of the system will equally well satisfy. In short, Kripke's argument entails that there is no possible world in which a definite and unique pure function or FSA supervenes *intrinsically* on the physical properties of the physical system. So both realization physicalism and supervenience of the formal on the physical are undermined by Kripke's argument.

Because supervenience cannot hold between formal properties and physical properties, it is natural to conclude that in any two possible worlds in which Γ is taken to compute different functions, this will be due not to any intrinsic physical properties of Γ (which we leave unchanged in both possible worlds) but to something extrinsic to it, such as the divergent intentions of different designers or observers, different formal practices, etc. The disjunction can only be true of the physical system by extrinsic denomination and thus in relation to something other than the physical structure and configuration of the mechanism. The physical structure and configuration might be necessary, but it cannot be sufficient for the realization of a pure formal structure. Functionalists need not hold on to supervenience, of course. But realization or implementation physicalism fares no better, and type physicalism about formal rules, functions, and operations is out of the question.

2. LANDE ON UNDERDETERMINATION VS. INDETERMINACY

Despite the fact that Lande's argument against functionalism (Lande 2011, p. 5, pp. 7-8) is so congenial to our Kripkean argument, he seems to go on to dismiss the formal indeterminacy of the physical as a plausible conclusion. He rejects that the formal indeterminacy of the physical is metaphysical or real as opposed to merely epistemological. He characterizes our conclusion as one of throwing "away any case of underdetermination as a case where the physical system doesn't realize anything at all. The antecedent of (UR), that Γ is a realizer, is false" (Lande, 2011, p. 5). By (UR), Lande means *his* principle of Unique Realization, which he states as

> *Unique Realization* (UR). For any physical realizer *P*, all the facts about *P* entail that it realizes an FSA *f*, and for every *f** that *P* realizes, *f** = *f*.

Though (UR) is fully compatible with (DR), it is much weaker. Lande takes a case like Γ's to constitute a counter-example to (UR), which is not a definition of realization but rather a statement about what the physical facts must entail if the physical realizer is to be said to realize any function. The difference is not clear. Lande seems to be arguing for the claim that none of the physical properties of Γ, not even under a maximally specific physical description, entail one function over another and therefore cannot sensibly be said to allow anyone to determine which function Γ computes. The result is still devastating for any theory that purports to identify formal reasoning (and human mental phenomena in general) with abstract structures, properties, and functions that supervene on or are realized by physical structures, processes, and properties. But it hasn't the same weight as (DR), as it leads to an epistemological rather than an ontological conclusion. It is, I must submit, not clear what Lande is trying to do here and it is far from obvious how he can hold that his argument doesn't pass ontological muster.[166] But let us consider his reasons.

Lande acknowledges (though not in these terms) that the Kripkean line of reasoning that we've developed can justify building (UR) into a definition of realization, such that

> 05) A physical system realizes an FSA if and only if it realizes that FSA and no other.

Notice that 05) is ambiguous. If it is to be read as equivalent (DR), then the qualification 'at the same time and in the same respect' should be read into it. At any rate, Lande is hesitant to build something like (UR) into the definition of realization and countenance something along the lines of (DR). Why? Well, it is not immediately clear. Lande would surely agree that our argument shows that which function or FSA a physical system, e.g. Γ, realizes is physically underdetermined. But he resists the conclusion that therefore a physical system, e.g. Γ, realizes no function or FSA at all. Here are his reasons:

> Any case of underdetermination, then, wouldn't be a case of realization. But this seems too restrictive. Any suitably complex analogue of [e.g. *Γ*] will, on this definition, not

[166] Note that functionalism explains the human mind in general, not only higher intellectual activity, in terms of the physical realizations of abstract functions, roles, or functional networks. Hence, functionalism if true would entail, on account of Kripke's argument, that the human mind is epistemically indeterminate and inaccessible. That is, we cannot know whether we think, understand, reason, calculate, infer, etc.! In fact, it is not merely that we cannot know which mental features are true of each of us, but that we cannot even know whether we have mental features or a mental life at all! This is absurd. The problem is to show why it is absurd. We will come back to this is chapter 4.

> count as a realizer. But it seems that we could extend examples of underdetermination to cover ordinary desktop computers and even human brains. So, if we rule out all cases of underdetermination, we seem to overly restrict what can count as a realizer. In any case, even if the functionalist should desire the truth of (UR), it is not clear what would a priori justify building (UR) into the very definition of realization, except as an ad hoc defense.
>
> (Lande, 2011, p. 7)

It is not clear how building something like (UR) into the very definition of realization could constitute a defense of any kind, ad hoc or otherwise, in favor of functionalism. Bet that as it may, our response to Lande's qualms is clear: what justifies building (UR) into a definition of realization is that there cannot be such a thing as the realization of a pure (mathematical or logical) function without satisfying (UR). Indeed, it is a minimal requirement for realizing a pure function that any given genuine case or application of such a function be determinately a case of *that* function and not also of an incompossible one (otherwise, it would not be a genuine or determinate case of *that* function).[167] But if none of the physical facts and features relevant for the realization of a pure function entail and determine that it is one function rather than another that is realized, then they entail and determine no unique or definite function and hence no function at all. For as we have seen, it cannot be true of a genuine case or application of, say, addition that what one is doing when one is genuinely adding is the same thing that, potentially and counterfactually, would *not* give you a sum (e.g. '5' for numbers greater than '57') were other numbers being calculated. To hold otherwise would be to make incompossible, i.e. contradictory, answers true at the same time and in the same respect. The rule of addition or identity is dispositive for any of its possible cases, and thus determines the formal correctness of any possible (i.e. actual, potential, and counterfactual) case or application. A formal rule excludes any other incompossible rule from being applied in the same sense whenever it is being applied. Furthermore, the fact that formal indeterminacy applies to *any* physical system, including desktops and brains, *is* precisely the point, and worries about it being overly restrictive are irrelevant if formal determinacy is required for the real, objective realization or implementation of a pure function. Indeed, what seems *ad hoc* is to block the argument from leading to where it naturally, and logically, leads.

Lande's qualms about inferring an ontological conclusion on seems to be motivated on the one hand by pragmatic worries and on the other by epistemological constraints. This latter constraint makes Lande's objection a variant of what I have called the 'epistemological objection'. This objection will be dealt with in full in chapter 5. Roughly, the objection claims that Kripke is confusing epistemological issues (formal underdetermination) with ontological issues (formal indeterminacy), and thus fallaciously reasoning from the epistemic inability to read off determinately which function a physical system realizes to the ontological impossibility of the physical realizing a definite pure function. From the fact that it cannot be known (i.e., be epistemically determined)

---

[167] Isn't this begging the question? The problem is that it is not clear what Lande's issue with formal indeterminacy in the ontological and the metaphysical sense is. If what Lande purports to suggest is merely that the physical properties of a computing machine do not, not even under a maximally specific physical description, suffice *for us to read off* a unique, definite function while leaving it open whether the physical system does indeed compute a determinate and unique function, then his argument is merely epistemological and can have no metaphysical consequences. Yet it does, as his appeal to a 'maximally specific physical description' of the physical system shows. It seems to me that Lande is just confused as to what follows from his argument. At any rate, an epistemological version of the argument would still be devastating for functionalism, computationalism, and token-physicalism in general about mental and psychological phenomena since their physical realizations would be mentally or psychologically underdetermined. Take formal understanding. If it were wholly physical, it would be formally underdetermined. But it isn't. We know and do mathematics, logic, formal linguistics, etc. Therefore, our formal understanding cannot be (wholly) physical. We will deal with again in chapter 5.

which function a device like Γ realizes it does not therefore follow that it does not or cannot realize any. Similar objections have been raised against Kripke's Wittgenstein, arguing that Kripke equivocates between the epistemic and ontological meanings of 'determine'. This objection, however, would be at odds with Lande's own argument. For Lande's (UR) says something *about* the physical realizer, not about the epistemic abilities of any potential observer. But since Lande seems to suggest that his argument is an underdetermination argument and should not be taken to imply 'no realization' (i.e. ontological formal indeterminacy), I take his objection to someone making an inference to ontological and metaphysical truths on the basis of a Kripke-style argument to be motivated by epistemological and pragmatic concerns.

Two possible and non-mutually exclusive responses would be to say that (1) the anti-functionalist argument grants the observer maximal epistemic access to the physical facts, properties, and processes of the computing system in question, and only to those relevant for the realization of the given function (cf. Quine, Miller, Kripke), and that (2) the Kripkean problem concerns whether there is any *kind* of physical feature, property, and process that could intrinsically determine a definite pure function, *not* whether we have the required epistemic abilities to determine whether such a physical thing actually exists (cf. Kusch). As far as (1) is concerned, the epistemological objection betrays a misunderstanding of Kripke's argument. For the claim being made is that even granting idealized epistemic access to some observer, i.e. access to all the relevant physical features in a computing system for the application of a given pure function, no such function could still be said to be determinately realized by the physical computing system. Hence, Kripke's remark (Kripke, 1982) that not even an omniscient being would be able to determine which function is being realized, because none is.[168] This ties neatly with (2), for if there cannot be such a thing as a physical fact, process, or feature which physically determines a unique pure function, there cannot be knowledge of any such thing, for there is nothing to be known.[169] There can be, however, knowledge of the indeterminacy of the physical with respect to pure forms. This is what we have. The reason why we cannot read off any determinate pure function from the physical structure of a physical mechanism is because no pure function is being carried out by the physical mechanism. In the next

---

[168] "Note also that the meaning scepticism which is the intended upshot of Quine's argument is *constitutive scepticism* (scepticism about the existence of a certain sort of fact), as opposed to traditional *epistemological scepticism* (which concedes that the sort of fact in question exists, and then questions our right to claim know- ledge of that sort of fact). It is important to bear this distinction in mind, because it would be easy to make the mistake of seeing Quine as propounding only epistemological scepticism about meaning. Easy, because his argument has a distinctively epistemological flavour: given all of the facts about stimulus meanings, you are still not in a position to know which of the competing translation manuals is the correct one. But note that although Quine's argument has this epistemological flavour, it questions whether we could know which manual was correct, even given *idealized epistemological access* to all of the *possible* facts about stimulus meanings and the natives' behavioral dispositions. Even if we imagine our knowledge-acquiring powers with respect to stimulus meanings idealized, we *still* would not be in a position to justify our choice of one of the competing translation manuals over its rivals. It is this idealization which allows Quine to argue for a constitutive scepticism *via* an epistemological argument." (Miller, 2007, p. 145)

[169] Perhaps it would be more precise here to speak of the issue being epistemically determinate in the sense that it *can* be known that the physical cannot realize a unique and definite function. We have shown this much. In consequence, we can say that it is epistemically determinate that the physical is, with respect to formal features, ontologically indeterminate. But "ontological determinacy" and "epistemic determinacy" (underdetermination) are not symmetrical and can be used in slightly but important different ways. Suppose an objector of ours insists on using "epistemic indeterminacy" and "underdetermination" in such a way that anything that is ontologically indeterminate with respect to two or more pure functions will be underdetermined (epistemically indeterminate) with respect to those functions in the sense that it will be undecidable whether any (and if any, which) function is being physically realized. On this view, formal ontological indeterminacy implies formal underdetermination, though not the other way around. In chapter II we will show that human formal understanding is not formally underdetermined, and thus cannot be ontologically indeterminate in form.

two chapters, we will come back to the epistemological objection and consider it in relation to Ross's argument.

3. STABLER'S WAY OUT: THE CAUSAL-COUNTERFACTUAL ACCOUNT

One natural way to try to dissolve Kripke's problem is to build into the definition of concrete realization or implementation the requirement of proper functioning or conditions of normal operation. In the last section, we saw one such attempt in Stabler's definition of what it is for a physical object to realize a mathematical function. "We require first that the states of the system can be interpreted as representing the elements of the domain and range of the function, and we require that (in certain circumstances) if the system is in a state representing an element of the domain of the function, physical laws guarantee that it will go into a state representing the corresponding element of the range of the function." (Stabler, 1987, p. 3) According to Stabler, a physical system computes a function *f* if, and only if,

(1) there is an interpretation function *In* which maps a set of finite sequences of physical 'input' states of the system onto the domain of *f*, and an interpretation function *Out* which maps a set of finite sequences of physical 'output' states onto the range of *f*, such that
(2) physical laws guarantee that (in certain circumstances *C*, and *if the system satisfied conditions of normal operation N for long enough*), if the system went successively through the states of an input sequence seq_i, it would go successively through the states of the corresponding output sequence seq_o where $\text{Out}(seq_i) = f(In(seq_i))$.[170]

There are several important qualifications here that need to be fleshed out. First is the inclusion of certain specifiable circumstances *C*, which must obtain if the physical system is to even begin to compute the relevant function (say, that certain buttons be pressed, certain switches be hit, or what have you). Second is the claim that physical laws guarantee that a set of unique physical conditions (i.e. *C* and *N*) for the physical mechanism will obtain. Thus, in the case of Γ, we can say that physical laws guarantee that that if a voltage pulse of such-and-such magnitude and duration is administered to Γ at its input point, then a voltage pulse of such-and-such magnitude and duration will be transmitted by the conductor to its output terminal in a finite amount of time. Physical laws guarantee that in the case of Γ this set of unique physical conditions will (ceteris paribus) obtain (Cf. Buechner, 2011, pp. 362-3). Stabler takes these conditions to be the ones necessary for the computation of the infinite identity function by Γ.

This leads us to the third condition, that the physical system must satisfy *conditions of normal operation* (Stabler also referred to these conditions as "background conditions of operation"). These conditions are meant to be physical. The requirement is necessary because break downs, malfunctions, and decay will happen eventually. But if no malfunction happens, such that system

---

[170] An alternative formulation prefers "physical states" over "sequences of physical states":

(1) there is an interpretation function *In* which maps a set of physical 'input' states of the system onto the domain of *f*, and an interpretation function *Out* which maps a set of physical 'output' states onto the range of *f*, such that
(2) physical laws guarantee that (in certain circumstances *C*, and *if the system satisfied conditions of normal operation N for long enough*), if the system is in a state *i* then the system will go into state *o* such that where $\text{Out}(o) = f(In(i))$.

We've chosen the specification in terms of sequences of physical states rather than physical states. See (Stabler, 1987, pp. 3–9) for the justification behind this choice.

continues to satisfy its normal conditions of operation, it will compute arbitrary values of whichever function it was computing. Since, however, malfunctions and breakdowns will surely happen, we must cash out the system's true computational description in terms of *counterfactuals*, that if the physical computing mechanism were not to malfunction or break down, it would realize whatever computation it was supposed to realize. Take Γ. The appropriate computational claim about Γ is supposed to be settled counterfactually: if the wire in Γ had not decayed after two voltage pulses and thus had continued to satisfy normal conditions of operation, physical laws guarantee that it would have computed arbitrary values of the identity function.[171] The idea is simple and offers a rather straightforward solution to the problem: if the system is working properly and therefore satisfies the physical conditions necessary for computing whichever function it is supposed to compute, one can see and read off easily from the unique physical conditions of the system the function it computes. The physics of the device guarantee what the system will do. Therefore, we do not need an instruction manual, nor knowledge of the intentions of a designer, to know which function a physical mechanism computes. To put it in Stabler's own words,

> There is a natural response to [Kripke's] problem. It is just to point out that *if the system had continued to satisfy conditions of normal operation for long enough, it would have computed arbitrary values of the identity function*. In the case of the wire, the identity function is distinguished in this respect from other functions which agree only on actually performed computations; the physics of simple circuits tells us this. For example, we do not want to say that the wire realized a function f' such that $f'(58) = 5$ [or, $f'(3) =$ null] because we know from the simple physics of the device that if it had continued to satisfy the background conditions (of being a simple conductor), it would *not* have computed this value.
>
> (Stabler, 1987, p. 9)

Now, Stabler (Stabler, 1987, p. 14; cf. Buechner, 2011, p. 363) argues that the proper way in which to express a computational claim is not

(a) Physical computing system S computes function f,

but rather

---

[171] This is similar to Chalmers's approach to computational implementation. Both Chalmers and Stabler place counterfactual constraints on the simple mapping account. Chalmers explicitly adds a *causal* constraint: for any concrete computational transition of the form $s_1 \rightarrow s_2$, if the system were in a physical state (say, s_1) that maps onto S_1, it would *cause* the system to go into the physical state (say, s_2) that maps onto S_2. There's also a further constraint placed by Chalmers, namely a *localizational* constraint. For details, see (Chalmers, 1994, 1995a, 1996, 2011, 2011) There are other forms of computational description that do not build on a simple mapping account. An alternative to the varieties of simple mapping accounts is what is called 'the semantic account of computation' (see e.g. Fodor, 1981, 2008). The semantic account takes it to be the case that there are physical states that are, intrinsically and non-relationally, representations. (There is, of course, no consensus on what makes a physical state a representation.) Though we do not concern ourselves here with semantic accounts of computation, it should become readily obvious when we deal with the semantic corollary of Kripke's argument that the physical is equally semantically indeterminate. Kripke's argument (or, rather, our adaptation of Kripke's argument) likewise shows that no material property of any physical state can determine a unique or definite formal content. Physical representations are physical objects that have content *denominatively*, i.e. in relation to something extrinsic to them (e.g., human linguistic conventions). No physical process, feature, or mechanism can solely on account of its physicality or materiality count as a representation of anything formally relevant—at least if by 'representation' we mean conceptual content represented by certain symbols or images. In that sense, there can be no semantics in physical nature, only in language and linguistic understanding, and for this too matter is insufficient.

(b) (Given interpretation *I*, circumstances *C*, and normal conditions of operation *N*) physical system *S* computes *f*.

If *S* breaks down or malfunctions, (a) will cease to be true, but (b) remains true. Malfunctions and breakdowns are excluded from (b) due to the exigency of normal conditions of operation *N* and circumstance *C* needing to be satisfied. The idea is that whenever *S* malfunctions or breaks down, *N* and/or *C* fail to obtain. But if *N* and *C* obtain then given interpretation *I*, *S* computes *f*. The inclusion of conditions *N* is supposed to block Kripke's argument and to constitute sufficient grounds for the formal determination of a physical computing mechanism. Actually, the claim Stabler makes is much stronger: if *N* and *C* obtain then given interpretation *I*, *P* computes *f* and only *f*. In the case where *S* malfunctions or breaks down, which function it was computing is expressed counterfactually: if *N* and *C* had continued to obtain for long enough then given interpretation *I*, *S* would've computed arbitrary values of *f* and no other incompossible function *f**.[172] So even after the physical system breaks down or malfunctions, such that (a) becomes false, (b) still expresses a true computational claim.

The fact that (a) has become false is due to either *C* or *N* (or both) no longer obtaining. But if *S* were to satisfy *N* (or had it continued to satisfy *N*) arbitrary values of the relevant function would be (or would've been) computed. And physical laws guarantee both that, if nothing goes wrong, *N* will obtain and that, if something goes wrong, that *N* would've obtained. "On the present account, the thing that distinguished the identity function from all of the other functions which agreed only on the actually performed computations on our wire was the truth of the counterfactual, that if the wire lasted would only last long enough, it would compute arbitrary values of the identity function." (Stabler, 1987, p. 17)

It is an ingenious reply. Unfortunately, it fails. It's a *petitio* (Buechner, 2011, pp. 364-6; cf. Kripke, 1982, p. 28). Consider the following schema. Let *U* = 'if a voltage pulse of such-and-such magnitude and duration is given at the input point, the conductor carries a voltage pulse of such-and-such magnitude and duration to the output point (guaranteed by the physics of Γ)'. Now consider the following formulation of Stabler's argument,

(A) Physical laws guarantee that a unique set of physical conditions *U* for *S* would've obtained had no malfunction or breakdown occurred.
(B) The unique set of physical conditions *U* constitute the normal conditions of operation *N*.[173]
(C) If *P* had continued to satisfy physical conditions *U*, *P* would've have continued to satisfy normal conditions of operation *N*.
(D) Therefore, if *P* had continued to satisfy normal conditions of operation *N*, *P* would've have computed f_N.

To Kripke's objection that a counterfactual account is inevitably circular, Stabler responds with something like (A)-(D). In addition, he adds:

---

[172] To be precise, all conditions (that is, *I*, *N*, and *C*) must obtain for *S* to compute *f*. But we will focus almost exclusively on the normal conditions of operation *N*, since *N* is the operative constraint in Stabler's refined account of computational implementation (physical computation) that is supposed to refute Kripke's argument. If *N* does not obtain, then *I* and *C* become irrelevant for the purposes of refuting Kripke.

[173] What if we were to say, instead, the *U determines* the conditions of normal operation, which in turn determine the formal operation that is being physically implemented? Then, given how *U* has been defined, we would beg the question against Kripke.

> We do not include in our background condition for the wire any specification of the infinite table of the identity function. The fact that the identity function is what the wire computes is determined by the fact that the background condition is the presence of a conductor, and by the fact of nature that conductors conduct voltage pulses in a regular way… If there is no problem with the claim that if the wire *remained undamaged* it would transmit arbitrarily many voltage pulses, then there is no trouble with the equally substantial claim that, under the specified interpretation and background conditions, it realized the identity function.
>
> (Stabler, 1987, p. 18, my italics)

What Stabler fails to see is that "conditions of normal operation *N*" stands for a specification of the function that is supposed to be carried out. And this is equally true for "conditions of abnormal operation A_N", that is, for malfunctions or deviations from *N*. The categorial difference between physical conditions and formal (normative) conditions is being conflated by Stabler. There is no doubt that if the wire remains undamaged the physical mechanism will continue to do something entirely *physical*, namely, conduct voltage pulses of arbitrary magnitude. But whether "remaining undamaged" is part of the normal conditions of operation is not something determined by the physical structure and doings of the device alone. For, according to a Quid designer, after a sequence of two voltage pulses a damaged wire is precisely what is part of the physical system's (say, Γ) conditions of normal operation. Stabler needs *U* to be sufficient for determining *N*. But this is precisely what is in question. Stabler just assumes that this is so, and so begs the question against Kripke.

According to the Quid designer, Stabler's counterfactual

06) If Γ had continued to satisfy normal conditions of operation *N* for long enough, Γ would've continued to compute arbitrary values of the identity function[174]

has a false consequent since the normal conditions of operations *N* for Γ are that it decay or breakdown after a sequence of two voltage pulses, and therefore it is not true of Γ that it would have continued to compute arbitrary values of the identity function because it was not computing the identity function at all. Therefore, according to the Quid designer, the consequent of 06) is false and must be negated. The reason why Stabler cannot see this is because in constructing Γ he has given himself access to the design of Γ and therefore to the intentions of the designer. The only counterfactual that the physics (i.e. *U*) of the device warrants is

07) If Γ had continued to satisfy *U* for long enough, Γ would've continued to conduct voltage pulses of arbitrary magnitude

But this just means that if Γ had continued to satisfy *U* it would've continued to satisfy *U*. This is true, but vacuously true. But this is not where the *petitio* lies, though it is close by. We can concede proposition 07). Our problem is with the suggestion that conditions *U* determine a definite formal operation, rule, or function. That's the thrust of Kripke's argument. The problem, therefore, with Stabler's reasoning from (A) to (D) is the step taken at (B). (One can still argue that even if conditions *U* obtain indefinitely and this *is* a malfunction, supposing it was designed for the quidentity function, then it can still be said the system computes the identity function. It was a

[174] Notice Stabler's qualification "long enough". How long though? How much longer would be *enough* to settle issues of formal determinacy? There is no such thing. The very same incompossibility problem would arise no matter for how long certain physical conditions of physical computing mechanism obtain.

wrongly or poorly engineered Quid-device, but a perfectly well engineered Id-device. We've dealt with this objection above, see our treatment of the conjunctive objection). A unique set of physical conditions *U* is compatible with incompossible conditions of normal operation N_f and $N_{f'}$.

What exactly is the content of "normal conditions of operation *N*" and what exactly determines the precise content of these conditions? Press home the point and it will become clear that one cannot specify what is normal, nor what is deviant, without prior knowledge of the norm to which the system adheres or is supposed to adhere. In Stabler's account, it becomes clear that "normal conditions of operation *N*" is a placeholder for "the physical conditions for computing the identity function":

(E) If *U* had continued to obtain for long enough, then *N* would've continued to obtain for Γ.
(F) If *N* had continued to obtain Γ, then Γ would've continued to compute arbitrary values of the identity function.
(G) Hence, it is determinate which function Γ computes because it is determinate which function Γ would've computed (or, rather, would've *continued* to compute) had it not malfunctioned.

Stabler's sleight of hand is obvious. The physics of Γ guarantee that if the conductor had not decayed, Γ would've continued to conduct. To deny this would be to deny the laws of electromagnetism. Agreed. But whether this counts as "abnormal conditions of operation" (malfunctioning, deviation, miscomputation) or as "normal conditions of operation" (proper functioning, correct computation) *is not something determined by the physics of Γ*. For, again, according to an Id designer, *U*'s not obtaining anymore after a sequence of two voltage pulses constitutes a break from the normal conditions of operation of Γ. But for a Quid designer, *U*'s not obtaining after a sequence of two voltage pulses constitutes *the* normal conditions of operation of Γ. For the Quid designer, decay is included in the normal conditions of operation *N*. If Γ had not decayed, then *U* would've continued to obtain. True. But then, says the Quid designer, Γ would've malfunctioned or would've manifested deviant behavior and thus would not have continued to satisfy conditions of operation *N*. For something to count as normal there has to be a norm (e.g. a rule or function) which that something is following; therefore, "normal conditions" is consequent upon, not prior, to the norm (function). Hence, the appeal to "normal conditions of operation" presupposes a definite norm. In our case, an appeal to the conditions of normal operation to deduce the function which the system would've computed had those conditions continued to obtain is to appeal to the function the system was computing or supposed to compute in order to determine which function it would've been computed. The circularity is brutal. In short, as Buechner himself notes following Kripke, the counterfactual:

(6) If Γ had continued to satisfy conditions of normal operation for long enough, it would've computed the identity function

is nothing but a placeholder for

(6′) If Γ had continued to compute the identity function for long enough, it would've computed the identity function.

A tautology, but vacuous. There is no intrinsically and objectively determinate link between *a set of unique physical conditions* and a *unique (definite) abstract function* and thus *a unique set of normal conditions of formal operation*.

The lesson to be learned is that one cannot appeal to the normal conditions of operations of a physical computing system to determine the function or FSA it implements without circularity. If you know what the physical device is supposed to be doing, then you obviously know what it would've done had it kept doing what it was doing, but if you don't know which function is supposed to be realized, you can't know what counts as normal (and what does not) in any computationally or formally relevant sense.[175] Again, what counts as normal (or deviant) depends on the norm that is to be followed, so if the norm is indeterminate, what is normal (or what is deviant) will likewise be indeterminate, and one cannot say what's normal (or deviant) without presupposing and squeezing in the norm that is to be realized or followed.

Let us illustrate this using the sort of computational claim that, according to Stabler, is the appropriate one. Consider again (b):

(b) (Given interpretation *I*, circumstances *C*, and normal conditions of operation *N*) physical system *S* computes *f*.

In order to apply (b) to any specific case, one must squeeze in the very function that (b) is supposed to determine. For example, in applying (b) to Γ, the physics of Γ leave it open whether

(b´) Given interpretation I_A, circumstances *C*, and normal conditions of operation N_A, Γ computes the identity function.

or,

(b´´) Given interpretation I_B, circumstances *C*, and normal conditions of operation N_B, Γ computes the two-counter function.[176]

is the one that applies to Γ whenever

(a´) Γ computes function *f*,

is said to be true. But, evidently, if (b) is just a placeholder for (b´), then surely whenever (a´) is true the proper computational claim will be (b´) and Γ can be said to compute the identity function. But this is precisely what is in question. If (b´) is the correct computational expression of Γ, then (a´) should be read in terms of (b´) such that f = *identity function*. Consequently, when (a´) is true Γ is a properly functioning identity-function computing FSA and when (a´) ceases to be true (say, after a sequence of two voltage pulses) Γ is a malfunctioning identity-function computing FSA. If, however, (b´´) is the correct computational expression of Γ, then when (a´) is true Γ is a properly functioning two-counter-function computing FSA and does not cease to be one after a sequence of two voltage pulses. So Γ would be a properly functioning two-counter-function computing FSA and f = *two-counter function*. On (b´´), proposition (a´) would cease to be true only if the wire did not decay after a sequence of two voltage pulses. An Id designer would take the conductor without decay as satisfying

[175] That we generally do not bother with such issues in everyday life when we buy a regular calculator or use calculator apps is because the *intended purpose* of those devices or apps has become part of the shared, social, epistemic background conditions for interpreting their behavior. We generally know what they are doing and what they are supposed to be doing because we know, by shared social background knowledge, what they were designed to do.

[176] Recall our Interpretation A and Interpretation B, last chapter. We are taking it for granted that interpretations A and B include an interpretation or realization function mapping the input and output states of each to the domain and range of the identity function and the two-counter function, respectively.

normal conditions of operation, whereas a Quid designer would take it as a malfunction or a deviation.[177] The values computed thereafter would be correct for the Id designer, whereas they would be a mistake and incorrect for the Quid designer.[178] The exact same physical laws, processes, and mechanisms. This is a slightly different proof of the claim that formal laws are *not* physical laws, and of the claim that anything physical only imperfectly realizes anything formal.[179]

As has already been noted, the problem with the causal-counterfactual solution is that the antecedent of the counterfactual conditional is an implicit statement about the definite function that is in question, thus making it circular, vacuously true, and wholly uninformative. But it is important to point out that Stabler needs to engage in a good deal of abstraction and idealization (e.g. counterfactual idealizations) when considering the formal features attributable to Γ. For, given our argument, Γ (or any other physical system for that matter) would need to continue to satisfy *N* indefinitely, since the Kripkean problem would emerge no matter at which point in the operating system's lifetime a decay or breakdown occurs. But an idealization of a physical system is *not* a physical description of the actual system and as such attributes to the actual physical system properties or features it cannot possibly have such as

---

[177] Let me also note that the Id designer or interpreter need not take the wire's decay as a malfunction; but rather as a physical limitation. Even physical computing mechanisms, for all their prowess, face limitations on account of just being physical. Some things will just wear out. But this is to be expected, and the Id designer can take this to be a mere physical limitation on the implementation of the identity function, or of an identity computing FSA. Indeterminacy issues will nevertheless abound, as the decay would not be taken as a limitation to the formal operation of computing the two-counter or quidentity function.

[178] For the Id designer, the device would continue to compute arbitrary values of the identity function. For the Quid, the device would be incorrectly computing the two-counter function. But why is saying that the system now incorrectly computes the two-counter function not be compatible with saying that it correctly computes the values of the identity function? Why can't the Quid designer now say, after the system fails to decay after the second voltage pulse, "the physical system is now an identity computing device, and no longer a two-counter computing device"? Note that the Id designer would have to say the inverse in the case where the wire does decay, such that if the wire decays or had decayed, it would or would've computed arbitrary values of a two-counter function. This now leads to absurdity. For, as we have seen, it would lead to contradictory counterfactual claims. According to the Quid option, if Γ had not decayed, it would've computed arbitrary values of the two-counter function. According to the Id option, if Γ had not decayed, it would've computed arbitrary values of the identity function, that is, it would not have computed the two-counter function. This is a *reductio*. Recall also that, even if decay happens after the second voltage pulse, there's an indefinite number of pure functions incompossible with the two-counter function that the finite behavior of Γ would still satisfy. Secondly, on the aforementioned reasoning it turns out that if the wire does decay, the system computes the two-counter function. But if the wire does not decay, the system computes the identity function. But this turns the computation of function into an *accident* and dependent on such aleatory events as decay, break down, and so forth. But one does not follow (i.e. genuinely and really follow) rules accidentally. One can stop at a red light accidentally and by coincidence, because the car engine shot down, the tires went flat, or because a cramp in my leg placed my feet right on the brake pedal. But none of this is hardly a case of having *followed* the relevant traffic rule. It is a case of acting or behaving as if one had followed a rule. In a loose sense one can say that it is behavior that "accords (or can be made out to accord) with the rule" without constituting a real case of that rule. Thirdly, this view would seem to imply that there are no such things as mistakes or malfunctions, for any apparent deviant behavior can be made out to accord with another function for which the behavior is not deviant. But then it makes no sense to say of a physical thing that it makes mistakes or that it carries on incorrectly. Actually, this seems to concede our argument. For given that what counts as normal and correct (and deviant and incorrect) depends on the norm or function being followed, there cannot be any such thing as "formal mistakes" in the physical, except in an observer-dependent way. Be that as it may, the response ultimately succumbs to a *reductio ad absurdum*, as we've already pointed out, and cannot offer a non-circular way of determining what Γ is *really* doing, whether prior to, or after, or in the absence of, decay. These objections were already dealt with above (see our argument on §1 against the conjunctivism and disjunctivism about physical computation) and were shown to be unsuccessful.

[179] So, contra Stabler, the infinite behavior of a machine (under some particular interpretation *I*, in circumstances *C*, and with respect to background conditions *N*) is *not* completely determined by the physics of the machine.

1) no break downs, decay, corruption, etc.,
2) no malfunctions or deviant behavior, and
3) infinite lifespan and computational (e.g. memory) capabilities.

Yet all of this, again, just concedes the argument: that the actual (and potential) physical properties of the actual physical system do not suffice to determine a definite or unique pure function. And since how we idealize a physical system depends on which function we take it to compute, idealizations cannot be non-circularly used to dissolve Kripke's objection. Appeal to abstractions and idealizations to cash out, for example, the computational counterfactuals that are true of a given physical mechanism will be hopelessly circular and thus fail a response to Kripke's argument.

Now, in his critique of Stabler, Buechner concedes that if both 1) and 3) could be true of a physical system, say Γ, then Γ would compute the identity function and Kripke's argument would not apply.[180] But this is not right. For not even an actual infinity of computations without decay or malfunction would suffice, especially for the computation of the values of infinite functions.[181] For an actual infinity of computations cannot produce both what it actually computes and what it would've computed had the functors been otherwise. An actual infinity cannot exhaust all the *possible* cases, and that is all we need to make our case. Suppose we define a function, called skidentity, which yields the same values as the identity function for any natural number, but diverges from it for any negative integer and irrational number. Since the set of natural numbers is infinite, it would take a physical system like Γ an infinity of actual computations just to compute the identity function over the natural numbers! But then it would be indeterminate whether Γ's infinitely many performed computations are a case of identity or skidentity, since skidentity would yield the very same values as identity for the very same actual infinity of performed computations over the natural numbers. The same would apply if we defined a function which differed from the identity function only for any

[180] "If we lived in a world in which (i) malfunctions of any kind cannot occur, (ii) PCMs last forever—a world without error of any kind, (iii) we knew this to be so, and (iv) we were immortal, the Kripke's problem for functionalism would not arise." (Buechner, 2011, p. 353) The problem with this statement is twofold. First, the notions 'malfunction' and 'error' are ambiguous. The concepts of 'malfunction' and 'error' always presuppose a certain norm or standard to which something is not adhering. Some standards or norms seem to be purely natural, and therefore not univocal with but analogous to our notion of 'rules'. This is readily obvious in non-artificial material phenomena like living organisms (e.g. liver failure, damaged blood vessels due to retinal hemorrhage, etc.). We can talk of 'natural' or 'organic' functions here. These malfunctions are *not* observer-dependent. But in cases of physical computing mechanisms that are artificial, whether a physical process counts as a malfunction will depend on what is meant by 'malfunction' or 'error'. If it means 'miscomputation', 'formally deviant behavior', 'wrongly applying or misapplying a formal rule', then Kripke's argument arises. If we mean something like 'the wire decayed', 'the machine broke down', etc., then then the concepts of 'error' or 'malfunction' might be misleading in this context since these are normative concepts and therefore point to deviation from some standard of correctness. But we show that no *formal* standard of correctness can exist in, be realized by, or supervene on anything physical. Now, in cases in which we can talk about physical malfunction and error (e.g. neurological malfunctioning), it must be clear that natural functions cannot be identified with formal or pure functions, for the former will be formally indeterminate and the latter will physically indeterminate (see next chapter, the transcendent determinacy of the physical). Formal errors cannot be reduced nor accounted for fully in terms of organic or neurological errors and vice versa. Though, of course, neurological impediments or injuries can and do affect the acquisition, retention, and exercise of formal understanding. Second, conditions (iii) and (iv) seem superfluous. If the point is merely epistemological, then yes. But if it isn't, it is irrelevant. It is, however, relevant as an attack on functionalism since it would entail or require idealizations about us that are and cannot be true of us. So unless these conditions are true, we cannot know which functions and functional network our psychology is the physical realization of.

[181] I am assuming, *for the sake of argument*, that an actual infinity of anything makes sense and that it makes sense, in particular, of something physical. I do not think this is an intelligible notion, however, especially if predicated of physical things, but even if we grant it, we'll see that the 'physical' would not be in any better position.

ordered n-tuple of numbers ordered from greater to lesser—"backwards" so to speak. The Kripkean point is not restricted to functions sharing a *finite* segment of equal or "compatible" cases.

Let us finish up our objection to the causal-counterfactual account. Stabler was aiming to give a plausible account of how infinite functions, like the addition and the identity functions, may be intrinsically realized by finite physical computing systems and thus refute Kripke's argument. His solution was to build into the definition of "realization or computing a function" the conditions of normal operation, which would objectively determine which function the physical system realizes without the need to appeal to the intentions of the designer. But the specific content of "normal conditions of operation" is determined, and can only be specified, by the function the physical system realizes. Hence, Stabler's response is viciously circular and epistemically vacuous, as it entails that we can come to know which function a physical system computes only if we know which function the physical system computes. Indeed, the only reason Stabler can confidently take his counterfactual or idealization to be true and thus a reliable guide to which function a physical system computes, is because he has antecedent knowledge of the system's design, of what it was meant to do. In other words, antecedent knowledge of the intentions of the designer. We agree with Stabler that "computational claims must be relativized not only to particular interpretations and particular circumstances of computation, but also to particular background conditions." (Stabler, 1987, p. 10). But what Stabler failed to recognize is that for his argument to work the so-called background conditions must already contain the *formal* conditions in question, and which formal conditions are the ones applicable to the physical mechanism under consideration is precisely what is in question. There is both a physical counterfactual and a formal counterfactual of the physical device, and that the gap between them is non-existent is precisely what needs to be shown rather than assumed. Physical computations must be relativized to the intentions of a designer or interpretations of observers.[182]

---

[182] An alternative would be to severe the ties between physical conditions *U* and conditions of normal operation *N* in our counterfactual account, and still maintain that the physical conditions in themselves determine the proper conditions for that system's capacity to compute the identity function. This would give a counterfactual that will remain true of any physical system *S* even when the physical conditions in the antecedent are not conditions of normal operation. We could say, for example

01) If physical system *S* had satisfied physical conditions *U* for long enough, *S* would've computed arbitrary values of the identity function.

We are assuming that physical conditions *U* are indeed the sort of conditions which implement a certain computation. But the qualification "long enough" is unhelpful, for as long as the system's actual and potential operating life-time is finite, indeterminacy problems will abound. There is no such thing as "long enough" in the determinant conditions of the realization of infinite functions. But consider

02) If physical system *S* would continue or had continued to satisfy physical conditions *U* indefinitely, *S* would compute or would've computed arbitrary values of the identity function.

Can this be objectively, i.e. observer-independently, *true* of *S*? Only if the physical conditions *U* can determine a definite function. But we've shown that it cannot. Furthermore, as we have shown (below, main text) that not even an *actual* infinity of physical computations can avoid formal indeterminacy, for nothing can do both what it does and what it would've done instead, and with sufficient ingenuity one can construct a function that coincides with another function for, say, rational numbers or natural numbers computed in a certain order, but which differs for, say, complex numbers and quaternions or natural numbers computed in a different order. Be that as it may, physical conditions *U* do not suffice to determine a unique formal operation.

4. WHAT ABOUT SEMANTICS? SEMANTICS AND COMPUTATION

One objection that can be raised at this point concerns the scope of our argument. It could be said that we've only argued against causal, counterfactual, dispositionalist accounts of physical computation which are accounts that explain the physical implementation of an abstract, formal structure in terms of the former instantiating a suitable causal pattern or configuration that is isomorphic with or "mirrors" (Chalmers, 1996, 1994, 2011, 2012)[183] the abstract formal structure of the computation. We've argued that any such physical causal patterns and configurations are isomorphic with the formal structure of incompossible computations. So someone can object that our argument works only against, and is directly only at, non-semantic accounts of physical computation. All that follows from our Kripkean argument, the objector says, is that simple mapping accounts (Godfrey-Smith, 2009) and purely structuralist accounts (Rescorla, 2013) of physical computation are incorrect. Once we recognize that physical computation, of the sort relevant to an explanation of formal understanding and operations, is semantically determinate, that is, it's an effective procedure over representations and therefore meaningful symbols, we will see that it can be determinate which formal operation a physical computing mechanism is carrying out. The formal structure of a program or a computation does not suffice to fix a "unique semantic interpretation" (Rescorla, 2013, p. 688) and therefore physically implementing a definite formal structure requires more than isomorphism, more than the physical causally-structurally mirroring it.

To illustrate the objection, let us take a look at two ideal physical computing mechanisms that are causally and syntactically symmetrical or identical but semantically and therefore operationally asymmetrical and divergent. The example is taken from (Rescorla, 2013) who uses it as a counterexample to structuralist accounts of computation[184] (that is, accounts of computation that do not consider content or semantics as *essential* to computations in general and take it to be a matter of isomorphism between the physical causal structure of a mechanism and the formal structure of the computation).[185] Suppose we want a machine that can find, for any two positive integers *n* and *m*,

---

[183] "A physical system implements a given computation when the causal structure of the physical system mirrors the formal structure of the computation … A physical system implements a given computation when there exists a grouping of physical states of the system into state-types and a one-to-one mapping from formal states of the computation to physical state-types, such that formal states related by an abstract state-transition relation are mapped onto physical state-types related by a corresponding causal state-transition function."(Chalmers, 1994, p. 392) This formulation is amended later on, but the underlying idea of isomorphism or 'mirroring' as that which bridges the gap between the physical and the formal remains throughout. "To implement a computation is just to have a set of components that interact causally according to a certain pattern. The nature of the components does not matter, and nor does the way that the causal links between components are implemented; all that matters is the pattern of causal organization of the system."(Chalmers, 1994, p. 401) Amongst the proponents of the isomorphic or 'structuralist' view, as Rescorla calls it (Rescorla, 2013), of physical computation are (Copeland, 1996, 1997; Dresner, 2010; Egan, 1995, 1992; Godfrey-Smith, 2009; Kim, 2011; Scheutz, 2002, 2001). Stabler's causal-counterfactual account of computation is in part a structuralist or simple mapping one but the semantic component given by the relevant interpretation seems to be one of the essential constituents of a computation (Stabler, 1987, but see 1986 for a non-representational account of linguistic behavior).

[184] Rescorla is not arguing for the opposite position, i.e. that content is essential for computation. Note that by content or semantics it is mean 'representation(s)' and therefore something is content-involving if it represents something. Rescorla tries to meet both opposites mid way, acknowledging that some computations do not involve content essentially and need not be individuated by their content(s) but that on the other there are computations (of the sort computer science is interested in) that do involve content essentially and are individuated in part by their content. For a now famous statement of the view that there is no computation without representation, see (Fodor, 1998). For other proponents of the "no computation without representation" thesis, see (e.g. Crane, 1990; Ladyman, 2009; Sprevak, 2010)

[185] Notice that the isomorphism or mirroring relation can go both ways. Suppose we want a formal, mathematical model of some physical phenomena. Then the formal structure is a model of the physical phenomena in question if it mirrors

their greatest common divisor, where $n \leq m$.[186] Let us give the following effective rules or algorithm for finding the greatest common divisor of any pair of positive integers n and m:

E1. [Find remainder] Divide m by n and let r be the remainder. (We will have $0 \leq r < n$.)
E2. [Is it zero?] If $r = 0$, the algorithm terminates; n is the answer.
E3. [Interchange] Set $m \leftarrow n$, $n \leftarrow r$, and go back to step E1.

where "$m \leftarrow n$" and "$n \leftarrow r$" mean, respectively, "replace the value of m for the current value of n´" and "replace the value of n for the current value of r". Thus, suppose we begin at step E1 and let the values be $m = 20$, $n = 5$. Then upon dividing 20 by 5, we see that that $r = 0$. Given E2, the algorithm comes to an end and we have found our answer, namely n. The greatest common divisor of 20 and 5 is 5. But suppose at E1, $m = 36$, $n = 5$. Then upon dividing, $r = 1$. At E3, we let $m = 5$ and $n = 1$ and go back to E1. Upon dividing, we find that $r = 0$. The algorithm stops, $n = 1$ is the answer.[187] Then then common greatest divisor between 36 and 5 is 1 which means that 36 and 5 are coprime or relatively prime.

Rescorla makes three important observations. First, that a programming language can render Euclid's algorithm.[188] Second, that semantics is necessary for this. Third, that here can be two physical computing mechanisms whose causal configuration and patterns and syntactic manipulations are exactly the same yet do not compute or carry out the same formal operation. Let us consider the first. According to Rescorla, Euclid's algorithm (let us call it E-A) is rendered within Scheme, a dialect the programming language LISP (still used in computer science), in the following manner (Abelson et al., 1996, pp. 48–49):[189]

```
(define (gcd a b)
 (if (= b 0)
  a
   (gcd b (remainder a b))))
```

where '`gcd`' stands for 'greatest common divisor', a and b for any positive integers, and 'remainder' is taken as primitive (Rescorla, 2013, pp. 687–688). This is an iterative process for computing the greatest common divisor of any two positive integers. Notice that a and b have to be content-involving, that is, a and b must range over numerals which represent numbers (i.e. positive integers) if the physical computing machine is to carry out E-A. This is taken to be a condition for the physical implementation of E-A within Scheme:

> A machine to carry out this algorithm must keep track of two numbers, a and b, so let us assume that these numbers are stored in two registers with those names. The basic operations required are testing whether the contents of register `b` is `zero` and

---

the physical, causal structure of the physical phenomena in question (e.g. computer simulations and mathematical estimations of a Hurricane's potential trajectory, speed, development, etc.)

[186] The greatest common divisor ('gcd' or 'GCD') can, of course, be defined for three or more positive integers, say 'gcd(a, b, c,…)'.

[187] This will be the same for any m and n if m is a prime number (like 5) and n is 1.

[188] Part of the reason for favoring Rescorla's discussion is that it allows us to run our argument with respect to programming languages (e.g. LISP), rather than merely with respect to machine models (e.g. Turing machines), in a very simple and straightforward way. Our argument can be made using other semantic accounts of computation and other programs formulated in Scheme or any other programming language for that matter.

[189] Rescorla's discussion of Euclid's algorithm and its rendition in Scheme is taken from (Abelson et al., 1996)

> computing the remainder of the contents of register `a` divided by the contents of register `b`. On each cycle of the GCD algorithm, the contents of register `a` must be replaced by the contents of register `b`, and the contents of `b` must be replaced by the remainder of the old contents of `a` divided by the old contents of `b`.
>
> (Abelson et al., 1996, p. 493)

Therefore, the states of the mechanism corresponding to *a* and *b* must be content involving and therefore representational. To see how E-A can be carried out within Scheme, let *a* and *b* be equivalent to *m* and *n* in E-A, respectively. We can initiate a computation in Scheme by introducing the numerical inputs '`20`' and '`5`', for the numbers 20 and 5, as the current values of *a* and *b*, respectively:

```
(define (gcd 20 5)
```

which immediately yields the numerical output '`5`', which is the greatest common divisor of 20 and 5. We can initiate the computation using any other numerical input, say,

```
(define (gcd 36 5)
```

which eventually yields '`1`'; or take Rescorla's own example,

```
(define (gcd 115 20)
```

which eventually yields '`5`' too, and so on.

Now, suppose we have two doppelganger physical computing machines, M_{10} and M_{13}, that have the same local, intrinsic, physical properties and configuration.

> M_{10} is employed by a society that uses base-10 notation. M_{13} is employed by a society that uses base-13 notation. We may stipulate that M_{10} executes our Scheme program. This stipulation is legitimate, because our own society can easily build such a machine. M_{10} and M_{13} have the same 'causal structure isomorphism type', since they are intrinsic physical duplicates. Nevertheless, M_{13} does not execute our Scheme program. Not, at least, if Abelson *et al.* correctly state what it takes to execute the program. When M_{10} converts input numerals '`115`' and '`20`' into the output '`5`', it thereby calculates the greatest common divisor of the corresponding numbers. When M_{13} performs the same syntactic manipulations, it does not thereby calculate the greatest common divisor of the corresponding numbers. The base-13 denotation of '`5`' (namely, the number five) is not a divisor of the base-13 denotation of '`20`' (namely, the number twenty-six). M_{13} does not execute the correct arithmetical operations, so it does not execute the Scheme program.
>
> (Rescorla, 2013, pp. 687–688)[190]

[190] "To explain why M_{10} executes the Scheme program while M_{13} does not, we must cite semantic properties of machines states. States of M_{10} have base-10 semantic interpretation, while states of M_{13} base-13 semantic interpretation. This difference does not reduce to relevant structural facts about the respective machines. The patterns of causal interaction are the same in both cases. What differ are not the patterns but the states composing the patterns. A machine implements the Scheme program only if it instantiates various states (representing the number *m*) and performs various

According to Rescorla, the only way to explain the computational difference between M_{10} and M_{13} is by an appeal to their different semantic properties. There is causal-syntactical symmetry between M_{10} and M_{13} but semantic asymmetry. On the basis of the former, there is no operational different. But it is the latter that constitutes the difference in the formal operations that M_{10} and M_{13} carry out. So "implementing the program requires more than instantiating a causal structure that mirrors relevant formal structure." (Rescorla, 2013, p. 688)

Now, how is this a plausible response to Kripke's argument? The semantic objection must trade on the assumption that Kripke's argument concerns only the causal structure of the physical mechanism. But since physical computation is not just a matter of the implementation of a causal structure that is isomorphic to the relevant formal structure, it follows that cannot as it stands be taken to refute any account of physical computation of a formal structure. The semantic account offers an alternative that does not reduce physical computation or the physical implementation of a formal structure to an isomorphism or simple mapping between the causally relevant structural features of a mechanism and the formal structure in question. M_{10} and M_{13} might be formally indeterminate with respect to their causal structure, but not with respect to their semantic 'structure' or properties. Hence, the semantic account is not vulnerable to Kripke's indeterminacy argument. But this is too quick and betrays a misunderstanding of Kripke's argument.

What exactly makes M_{10} and M_{13} semantically determinate and thus operationally asymmetrical? The answer is simple: that M_{10} has a base-10 semantic interpretation and M_{13} has base-13 semantic interpretation. Nothing about M_{10} and M_{13} blocks the semantic interpretations from being interchanges such that M_{10} is given a base-13 semantic interpretation and M_{13} a base-10 semantic interpretation. But this precisely Kripke's point. Nothing about the physical features of a computing machine can determine a definite formal operation; one needs the intended interpretation and therefore the operation of the system is relative to the intentions of its designer or the linguistic community in which it is embedded. Kripke's problem is, however, cuts even deeper. It points to the fact that even M_{10} will satisfy incompossible base-10 semantic interpretations (say, a base-10 and a qase-10 semantic interpretation). One problem is, of course, that the strings *qua* physical items in a physical implementation of a program do not contain any semantic properties intrinsically and on account of any of its physical properties. Here's Rescorla's take on this issue:

> But how can executing a computer program require having certain representational properties, when the program itself is just a meaningless string of signs? Aren't M_{10} and M_{13} programmed with the same Scheme instructions? How can there be any significant difference between the computations executed by M_{10} and M_{13}? I respond that there is more to a program than meaningless signs. The signs have an intended interpretation as instructions to do something, just as a recipe has an intended interpretation as instructions to do something. The intended meaning of 'add salt' is that one should add salt. The intended meaning of our Scheme program is that the machine should perform certain arithmetical operations. M_{10} and M_{13} are programmed with the same strings, but those strings contain numerals that have different meanings within the respective linguistic practices in which M_{10} and M_{13} are embedded, so the strings have different intended meanings. Despite syntactic overlap, M_{10} and M_{13} do not execute the same instructions. To claim otherwise suggests a use–mention confusion between the

---

operations (dividing *m* by *n*) whose natures outstrip any relevant pattern of causal organization. Thus, M_{13} is a counter-example to structuralism." (Rescorla, 2013, p. 688)

> numerals composing a programming language and the numbers denoted by those numerals (as used by the surrounding linguistic community).
>
> (Rescorla, 2013, p. 689)

But this is precisely our point! There is a fact of the matter as to what a physical computing mechanism does, but it is not a matter of its matter alone or of any intrinsic properties of it. It requires extrinsic determination and therefore has meaning only by extrinsic denomination (e.g. explicit, extrinsic stipulation from its designers about what its strings denote) via an *intended* interpretation, and impossible without a community's agreement in symbolic transformations and representational conventions (e.g. Arabic numerals). The physical system has no ability to represent anything, it is rather endowed from without with representational or semantic properties by stipulation and interpretation—which are things we humans impose for suitable purposes on the mechanism and the strings of a programming language. As Kripke said, if M_{10} and M_{13} had fallen from the sky, there'd be no fact of the matter as to which operations each computes—and I mean this both with respect to which syntactic rules they follow and which operation they carry out, which can partly depend on the type of admissible variable component (i.e. semantics). But what counts as 'admissible variable component' is interpretation dependent and imposed from without by the programmers or designers in accord with general linguistic conventions or technical stipulation.

This last point needs to be hammered down because we are led to think that the syntactic manipulation is definite in the case of both M_{10} and M_{13}, and from this Rescorla goes on to say that M_{10} and M_{13} are syntactically the same. But this begs the question if what is meant is that both M_{10} and M_{13} follow and apply the very same syntactic rules (for a purely syntactic account of computation, see Stitch 1983). But the basis for saying this is that their actual, causal structure and finite performance with respect to the relevant sort of strings is of the same kind. Yet the actual and potential causal pattern and performance *of each machine* can be made out to accord with incompossible syntactic rules. One does not, again, need two machines, M_{10} and M_{13}, to make the point; one will suffice. For any syntactic rule R that M_{10} or M_{13} is said to follow or conform to, there is at least some other incompossible syntactic rule R' which coincides with R up to a certain point or with respect to a certain type of strings but diverges from it afterwards or otherwise. So syntactic indeterminacy, which is a species of formal indeterminacy, cannot be avoided by a physical computing mechanism. We do, however, hold that there is such a thing a physical computation but it is neither analogous with what we do when we compute, as the exercise of an ability, nor explicable in terms of the physical implementation of formal structures or pure functions. Later on (see §7) we consider the proper way to talk about computation in the physical realm and how it can be intelligibly said that there is a fact of the matter as to what a physical computing mechanism does, computationally and formally speaking. For now, we take Rescorla to have proven our point: only by stipulation and therefore given an intended interpretation can it be said of a physical mechanism that it computes a definite a pure function or that it carries out definite formal operation. This is computation by extrinsic denomination, and it's more a case of simulation or approximation to the real thing than a duplication of it. It is simply not univocal to what we do.

Nothing in the semantic account undermines what we've argued for thus far. On the contrary, it confirms it. So long as a semantic account of computation does not attempt to ground the determinants of the structure of the formal operation that is being carried out in terms of the physical properties and structure of the system, it not undermined by Kripke's argument (unless it claims that the extrinsic determinant of the formal operation being implement is something entirely physical, e.g. the physical environment, in which case Kripke's argument would arise again; but

Rescorla attempts no such thing, appealing instead to social and linguistic environments).[191] Rescorla's semantic account of computational implementation does not seek and actually rejects any such grounding. When Rescorla points out that the Scheme's designers are committed to a form of denotational semantics, that they intend their talk of semantic properties to be literally true, that they construe the program as having representational properties and that therefore any physical mechanism that implements Scheme must do so, in part, due to its having the relevant representational properties, he is not saying anything that contradicts Kripke's argument. For the relevant representational properties are stipulated from without and are fixed and held constant by the intentions (i.e. the intended interpretation) of the programmers and designers. But in themselves and on account merely of their physical properties, the relevant strings that are intended to count as symbols and therefore as meaningful or representations do not count as having any semantic determinacy. There are two levels here. One is at the level of programming and thus of programming languages. The other is at the level of implementing the programming languages in (disparate but suitable) physical mechanisms. As we've said, the programming language becomes determinate by explicit stipulation and/or by piggybacking on a community's natural linguistic conventions and agreements. As Rescorla himself concedes, "[a] programming language has no inherent meaning in itself" though programmers associate a programming language with an intended interpretation (Rescorla, 2013, p. 686). The formal (read: syntactic) structure of a programming language is, though in itself semantically indeterminate, semantically determinable by, say, explicit stipulation from the programmers and their intended interpretation. But the same goes for the syntactic determinacy of the programming language (that is, the precise syntactic rules that constitute it). At the level of the implementation, say M_{10} and M_{13}, the physical mechanisms in question are both syntactically and semantically indeterminate yet determinable in both respects by stipulation and relative to the specific intentions of their designers. This has been our point all along: that one needs to look outside the physical properties of the physical system itself to determine what it is capable of and what exactly it does in terms of rule-following and operations, syntactically and semantically; for, such properties do not suffice for the metaphysical determination of form and content. The operational determinacy given to the physical mechanism by the intended interpretation of its designer is a relative and observer (designer) dependent one, and therefore held constant and sustained from without.

> (1) It is possible for there to exist intrinsic physical duplicates M_{10} and M_{13}, the former appropriately embedded in a linguistic community that employs base-10 notation, the latter embedded in a linguistic community that employs base-13 notation. Since M_{10} and M_{13} are intrinsic physical duplicates, they have the same causal organization.
> (2) M_{10} manipulates numerals whose meanings are given by base-10 notation, and it executes appropriate syntactic manipulations, so it thereby implements the Euclidean algorithm Scheme program. M_{13} manipulates numerals whose meanings are given by base-13, so it does not implement the Euclidean algorithm Scheme program, even though it executes the same syntactic manipulations as M_{10}.
>
> (Rescorla, 2013, pp. 694–695)

[191] Whether these environments can be reduced to the physical is another issue. But if our characterziation of mixed formal structures (see chapter 1) is correct and our account of language as involving, or being a condition for the possibility of, formal operations is also not too far from the truth (see chapters 1 and 5) then neither social nor linguistic environments (which are inextricably intertwined, anyways) can be entirely or wholly physical.

Again, nothing intrinsic to both M_{10} and M_{13} impedes them from swapping places so as to be embedded in different linguistic communities and therefore count as semantically and operationally different, M_{13} operating on a base-10 semantics and M_{10} on a base-13 semantics. That's all we need for our argument. Whether physical computation must essentially involve semantic properties is not our battle.[192] If they do, then all well and good. If they don't, then all well and good. But either way the syntactic (and, if any, semantic determinacy) of any physical computing mechanism will be an observer-dependent fact.

So even if the full-blown (see e.g. Fodor, 1998; Ladyman, 2009; Sprevak, 2010) or the moderate (Rescorla, 2012, 2013, 2014, 2015) semantic account of computation is correct, Kripke's argument will apply just as well to these accounts. Kripke's problem does not concern which theory best explains physical computation, but rather criteria of intelligibility and possibility: in which sense can it be said, intelligibly, that a physical mechanism implements a given formal structure, follows a formal rule, and computes a pure function.[193] In §8 we will see that if functionalism or computationalism as a metaphysical thesis about the nature of formal understanding takes that sort of understanding to be a species of physical computation, then it will be subject to very same indeterminacy considerations and the designers themselves will be in their intended interpretations, stipulations, and definitions equally indeterminate; except by extrinsic stipulation and interpretation (from who?) there are no formally and semantically determinate acts of understanding on the part of the designers. One needs to look elsewhere (say, the intended interpretation and stipulations of the designer of the designer?). But then it seems that, amongst other things, the computationalist faces an infinite regress unless there is a determinant which contains formal and semantic determinacy in an inherent, underived, and observer-independent way. A full defense of the non-*ad hoc* claim that humans (i.e. human formal understanding) are such a determinant will be the subject of the next chapter.

---

[192] We are, however, of the opinion that our formal activity is in general both semantically and formally determinate and that formal understanding is sensitive to the variable component (the content) insofar as a formal structure is that which remains or would remain invariable through the indefinitely many transformations and permutations of its *admissible* variable component (e.g., one cannot add the addition function or the ocean). Having said that, our argument concerns primarily *formal* operations *qua* formal and therefore applies to our ability to instantiate formal structures in the absence of any semantics (e.g., syntactic rules over uninterpreted symbols).

[193] Interestingly, Rescorla briefly touches on Kripke's (Kripke, 1982) rule-following considerations in a footnote but dismisses them as having no relevant effect on his anti-structuralist argument. Rescorla thinks that Kripkestein's rule-following considerations concern only rules with an infinite scope, such as the infinite series generable on the basis of the rule 'add 2'. For, says Rescorla, his argument in terms of M_{10} and M_{13} can be made just as well in terms of register machines, R_{10} and R_{13}, to which the same philosophical considerations apply but whose formalism can be modified so as to assume that the machines have a large but finite memory capacity (the Abelson *et al.* register machine formalism assumes infinite discreet memory capacity on the part of the register machine, something practically and in principle impossible for any physical mechanism). "Given my focus on register machines with finite memory capacity, we can concentrate entirely on rules with finite scope. For instance, rather than considering the remainder operation as defined over all numbers, we can consider a modified operation that applies only to numbers below some upper bound determined by the register machine's finite memory capacity. We can enumerate how this operation applies to all numbers below the upper bound. The enumeration might take longer than any actual human lifetime, but it is possible in principle for an idealized human. In principle, then, we can explicitly enumerate what a physical machine must do in order to realize the Euclidean algorithm register machine (with restricted memory). No Kripkensteinian worry arises, because there is no need to extend the programming instructions to larger numbers beyond our explicit enumeration. It is easy to show that R_{10} executes the desired instructions but that R_{13} does not." (Rescorla, 2013, p. 697 n. 8) But this is misconceived. Kripke's argument can be made just as well with rules with finite-scope, since an enumeration of how a given operation f applies to all numbers below the specified upper bound can be made out to accord with an enumeration of how another incompossible operation g applies to all numbers below that upper bound. We have already made clear in the main text how M_{10} (and M_{13}) can be made out to accord with incompossible formal operations; *mutatis mutandis* for R_{10} (and for R_{13}).

5. THE SUI GENERIS PHYSICAL FACT OBJECTION

This is a physicalist variation of the *sui generis* response to the skeptical argument (see e.g. Kripke, 1982, p. 68). The response goes something like this:

***Sui Generis* Objection**. From the fact that no physical thing at any level can be said realize a pure function it doesn't follow that whatever realizes a pure function must be immaterial. Perhaps we are dealing with a *sui generis* kind of physical process, unlike any other. An unspecifiable, non-standard kind of physical process and property, non-reducible to other physical processes but physical nonetheless.

To this objection, we could simply say: fine! If one appeals to *sui generis* physical processes and properties to account for the realization of formal structures, we then have to ask what is the content of the term 'physical/material' in '*sui generis* physical or material processes and properties'. It cannot be 'physical' in the sense 'physical' ordinarily and scientifically means, for then it would not be *sui generis* and it would be subject to Kripke's argument. In other words, it would have to be a 'physical' thing that is not subject to any of the physical laws and properties of *any* physical thing, otherwise it would fall within the scope of Kripke's argument. But then it is hard to see how such a sui generis fact could be a *physical* fact, for it is no longer clear what it could possibly mean to call it *physical*. For talk about 'a sui generis physical thing or process' would seem to depend on either a change in the meaning of the term 'physical' (in which case the objection either fallaciously changes the subject or concedes the argument by using 'sui generis physical fact' for what is really an immaterial/non-physical fact) or a nonsensical use of the term 'physical' (in which case the objection is confused and expresses nothing, for it makes no sense). At any rate, the objection concedes that with respect to 'non-sui generis' physical or material phenomena, our argument stands. So the onus is on the objector to specify the physicality or materiality of such a *sui generis* physical fact and to show that it is not a nonsensical, equivocal, or vacuous use of that notion.

We could augment this objection by claiming that from the fact that certain phenomena are not reducible to, or explainable in terms of, lower-level physical phenomena it does not follow that they are not physical or material. For example, suppose biological phenomenal are not reducible to lower-level physical (non-biological) phenomena. Does it follow therefore that biological phenomena are *not* material? Of course not. Furthermore, even according to the Aristotelians, though consciousness is not reducible to any lower level physical phenomena, it does not follow that therefore consciousness is immaterial. For *irreducibility is not a mark of immateriality*. This much, I think, is true and correct. So, the question now is, why therefore should it follow from the fact that formal reasoning cannot be reduced to, nor explained in terms of, any lower-level physical phenomena that it is therefore *not* material? A proper answer to this objection belongs to the next chapter (and chapter 5). Here we will limit ourselves to a sketch.

First, notice that we have not argued that formal structures are irreducible to physical structures. Our argument is not an argument from irreducibility. We have indeed claimed, and will continue to do so, that the former are inexplicable in terms of the latter and in this sense irreducible to them. But we have not taken this to be the reason why any potential realizer of a pure function must be not be wholly material. We have argued, instead, that no physical thing (thus, nothing that satisfies at least the three conditions given in §3), can realize a pure function determinately, that is, without also realizing incompossible ones. But it is a condition for the *real* realization of a pure function that it be determinate, and thus that it exclude incompossible ones. Hence, no physical thing—whether lower-level or higher-level—can really and determinately realize a pure function

since nothing physical can meet a necessary condition for the realization of a pure function. Therefore, physical processes do not, for they cannot, be used to explain formal reasoning, since something's being a real case of a pure function cannot consist in the *physical* realization of that function. Physical things can only realize pure functions extrinsically and in an observer-dependent way. (Of course, one can likewise claim that formal reasoning is nothing but physical (e.g., neural) processes and consequently unable realize pure functions either. We will consider and refute this objection in chapter 4, §3 and §4.)

Second, the nature of the structural principles constitutive of formal reasoning gives way to further considerations. Given what *pure formal structures* are—namely, structures whose identity, completeness, and existence does not require, nor depend on, any materialization—, and given that the physical cannot realize, entail, or determine such structures, the physical cannot bring those structures into being either.[194] Now, if human understanding realizes them (chapter 4) in a real, intrinsic, and determinate way, and if human understanding can even create or construct them (at least some of them), then not only can human understanding not be wholly material but it could not have been brought about entirely by material processes. At any rate, the claim to be made here is, roughly, that on pain of absurdity such pure formal structures must be at least structural principles of human formal understanding. Otherwise, there could be no such thing as formal reasoning. But this entails that formal reasoning cannot be (wholly) material. In this chapter, however, we've simply argued that pure formal structures cannot be physically realized in any real and determinate way. They can only be approximated or simulated (even to an impressive degree) by anything physical. Of course, one could deny that there are any such pure formal structures. But this is indefensible, and easily shown to be false. All we need to do is to construct and define specific formal structures (say, Kripke frames or the rules for universal quantification), test them against physical structures for intelligibility conditions (e.g., try to uncover the quantum effect of Zermelo's axiom of choice or ask which physical science might be best equipped to study the physical properties of Łukasiewicz logics), and our job is done. Similarly, specify any set of functions of any branch of pure mathematics and we would've likewise proven our point.

6. THE SLIPPERY SLOPE OBJECTION

A more forceful challenge to the Kripkean argument can be posed by asking why the same considerations that apply to pure formal structures do not extend to both mixed and natural structures. It seems arbitrary, indeed *ad hoc*, to limit the scope of the argument to pure structures. Take for example the biological process of mitosis.[195] We can likewise formulate an incompossible biological process, 'schmitosis', which the real processes satisfy and which behaves like mitosis for a certain amount of cellular divisions up until 2026 but diverges afterwards (for a fuller treatment of the 'schmitosis' example, see chapter 5, §8) . What in the biological processes themselves determines which one is being realized? Who is to say that it is a case of mitosis rather than schmitosis? And so on for any physical (i.e. chemical, biological, geological, neurological, etc.) process. It will not do to

---

[194] Which sort of material or physical component is constitutive of a Euclidean point or line or algorithm? Which actual or potential natural science should study the purported physical properties of the formulae for π? How much mass does *modus tollens* have? What about categorical affirmative propositions, are they composed of organic or non-organic matter? What is the entropic state of second-order logic? What about the neural properties of the Pythagorean theorem? Can the rule of exponents in algebra be subject to chemical change? How is set theory subject to the laws of thermodynamics? Can any metabolic processes impinge on the validity of deductive inferences? None of these questions make sense, because they hinge on a category mistake. The intelligibility conditions track the categorical difference between formal and physical structures.

[195] See e.g. (Dillard, 2014, 2011)

respond by saying that it does not apply to mitosis and schmitosis because they are not pure functions. This would be a text-book example of a *petitio*, since it assumes what is in question.[196] The question is, why are the Kripkean considerations only limited to pure formal structures?

The answer to this question is straight forward: because natural structures contain the matter proper to them in their definition, that is, in their essence, and one cannot specify fully a natural structure nor can it be constituted independent and in the absence of its proper matter. This is partly because natural structure are structures constituted by materialization and a specific type of matter. Any characterization of a biological process that abstracts away for the proper matter of the process will constitute an abstraction and idealization of the process itself (viz., a potential formalization, in our sense) and would thus be subject equally well to the formal indeterminacy argument. 'Schmitosis' is just one example of an abstraction and an idealization of the biological process in question. 'Mitosis' is the name we give to the biological process, which we can of course formalize for heuristic and experimental purposes. But the formalization can never be identified with the biological process itself, except by a sort of 'baptism' (cf. Kripke's rigid designator theory) and thus by extrinsic denomination. Of course, one could retort by saying that this is just as question begging as Feser's response (see n. 196), since it amounts to saying that natural structures are not formal structures and therefore not subject to the same considerations. But there is no getting around this. One could try denying that pure formal structures have their identity, completeness, and existence independent of any type or token material property, but this is demonstrably confused (see, e.g., n. 194). Pure formal structures require no materialization—not even the possibility of materialization. That is why they are not the proper object of any natural science and are not subject to any physical laws and processes (e.g. there is no such thing and cannot be such a thing as a quantum level in Peano set theory nor atom decay and radioactive residue in S5 modal logic). Not so for real, natural structures. They require a specific kind of materialization and cannot be what they are (nor be) without their proper materialization. This makes them be grounded in matter in a way pure formal structures cannot be, and the structure's proper matter, which is essential to the kind of structure it is, plays a role in the determination of, or in the persistence and enabling of, the structure. Therefore, any physical or material thing cannot be indeterminate with respect to a material, natural structure, since what it is for something to be physical or material is for it to be *at least* structured in certain specific material ways—ways, of course, subject to physical laws and regularities, and so forth. No such thing obtains for pure formal structures, whose essence does not and cannot contain matter.[197] The question that remains is what exactly do these pure forms *structure* or *inform*, and we will answer this in chapter 4.

---

196 Feser falls precisely into this basic, logical mistake in a response to Dillard (see e.g., Feser, 2017, cf. 2013). Since Feser's argument is about semantic determinacy, he's response to this objection amounts is merely an appeal to the fact that 'mitosis' and 'schmitosis' do not involve semantic content and therefore cannot be said to raise indeterminacy that pure forms like 'addition' and 'quaddition' raise. But the question is precisely why should the considerations apply only to semantic contents if similar considerations can be theoretically developed for biological and, in general material, functions? In fact, as Feser is well aware, Goodman developed such considerations with respect to color expressions. Plus, don't we have the semantically determinate concepts of 'mitosis' and 'schmitosis'? Wouldn't, on Feser's on account, the biological process in question be indeterminate with respect to such concepts? For my critique of Feser's version of Ross argument, see chapters 4 and 5. Incidentally, see also (Kusch, 2006) for another example of someone who subsumes "following a rule" (the formal aspect) under "meaning a rule" (the semantic aspect) in his treatment of Kripkestein.

197 An additional point would be to say that because natural structures are not abstractions there is no role for the intellect in determining or constituting the essence of physical, chemical, and biological processes such as metabolism, mitosis, photosynthesis, etc. These are found or discovered, but not constituted and determined in their being by human intellectual activity. But this drags us into the philosophy of mathematics which is beyond our scope and into issues in the philosophy of science as well as a host of related issues that will have to wait until chapter 4.

To conclude, it makes no sense to say that a biological organism and process is indeterminate (ontologically and metaphysically speaking) with respect to incompossible biological structures, for then it would not be a biological organism or process at all. To talk of a biological thing or process is to talk of something that is biologically structured and as such biologically determinate. (The same goes for any other material kind of structure and process). If we say that a certain biological phenomenon is indeterminate with respect to incompossible biological forms, then in order to be saying something that makes sense this would have to be taken either as a statement of our current epistemic state (say, that we cannot yet determine the precise structure, properties, and so forth of a given biological phenomenon and thus cannot decide between competing proposals, as opposed to saying that there is no biological structure or property at all), or as evidence that our descriptions are idealized and obtuse abstractions, which the biological will only approximate.[198]

7. IS THERE PHYSICAL COMPUTATION AT ALL?

Though I've already said that Kripke's argument is *not* a triviality argument, one can still raise the question as to whether my assessment is correct. After all, we are arguing that there are no physical facts, properties, and processes determine which abstract formal (mathematical or logical) function or description a physical computing mechanism implements or satisfies. And if it comes down ultimately to the side of the observer and the agreed upon, intended, interpretation and if there is no objective fact in the physical system that could make a given interpretation or observation true or false, then it seems that triviality or vacuity follows. But this is incorrect. This objection trades on an ambiguity in the notion of 'objective' and 'objectivity'. When we say that nothing physical can objectively determine which (or even that a) mathematical or logical function is being computed we mean to deny *realism* about physical computation (i.e., the view that physical processes can compute mathematical and logical functions *intrinsically*, i.e. in a way that is wholly independent of human intentions, conventions, interpretations, etc.) together with the claim that our computations and applications of mathematical and logical functions and rules are instances, a species, of physical computation. We have not yet argued that physical computation cannot explain human computation, though this should have begun to become clear by now. We shall do so in the following two chapters. But we have argued for the claim that physical computation is, in its formal determination, human-dependent (e.g., observer-dependent, interpreter-dependent, designer-dependent, or what have you). Nothing in virtue of physical properties or features alone can be a sufficient condition for the formal determination of real mathematical and logical operations. So realism about physical computation is false, it misdescribes or misrepresents the nature of physical computation (see §8 below120). The most plausible alternative view, which is the one we have argued for, would be a non-realist view of physical computation.

Now, recall that a non-realist view of physical computation (as we've characterized this view) is the view that the physical can only compute mathematical and logical functions in a manner that is wholly or partially *extrinsic*, i.e. dependent of human intentions, conventions, interpretations, etc. The question is, is it wholly or partly so? We say partly, for we do not claim (though we do not argue for this claim here) that any physical thing satisfies the necessary conditions for counting as a physical computing mechanism. That is, we are not pancomputationalists. We do not gainsay that a physical

[198] More needs to be said. I will have more to say about this in the conclusion when we consider the role of mathematics in the natural sciences and scientific knowledge, but it will be more in terms of a problem to be addressed by further research rather than a solution to the problem itself (e.g., the problem of explaining the applicability of mathematics to material reality).

mechanism if suitably configured (for not anything physical computes, e.g. rocks don't compute) *can* be a computing mechanism and if not, it cannot—at least in any philosophically interesting sense. There are real constraints on what can count as a computing mechanism. Only certain physical architectures and configurations are relevant for being a computational mechanism. Computer science, to name one discipline, is not in any way undermined by our argument. In fact, our Kripkean argument began by conceding for the sake of argument that both triviality and pancomputationalist arguments had been overcome. Our point is this: physical computation of mathematical and logical functions is not and cannot be *wholly* mind-independent, and therefore there is no such thing as intrinsic, non-derived, mind-independent physical computation of mathematical and logical functions. (A corollary of our argument (to be argued for conclusively in the next chapter) is that physical computation cannot the same as human computation (for want of a better term) and human computation cannot be the same as physical computation.) Physical computation of the relevant sort (that is, the use of physical computation to "duplicate" and explain human computation, i.e. our mathematical or logical abilities) is a hybrid of properties and capacities that are wholly physically determined in an intrinsic way and of properties and capacities that are formally determined in an extrinsic way (say, by extrinsic denomination). For example, it is an objective fact about your computer is the kind of physical mechanism that can run a determinate program and compute definite formal rules. Yet this objective fact of the matter is constituted in part by the intentions of the designer(s) and programmer(s) which constitute a sufficient condition of, or the missing link for, the formal determination of the physical mechanism. Without knowledge of the intended design and interpreted program there is no telling which amongst incompossible designs and programs the physical system is meant to satisfy. It is because we know the intentions of the designer, which have become part of our shared background conditions for interpreting such devices (imagine Aristotle being transported into the present time and being handed a pocket calculator or an iPad), that we know that most pocket calculators add rather than quadd. But we do not know this by any physical property of the pocket calculator, since *qua* physical mechanism it satisfies equally well the quaddition function or some other incompossible function which the finite states and operations of the mechanism will equally well accord with. We know this by shared interpretative practices, conventions about what constitutes a suitable mathematical, logical, or linguistic symbol (e.g., programming languages) and what constitutes appropriate manipulations and transformations of those symbols so as to fix the referents of the symbols and the rules which the manipulations and transformations determinately conform with. So in a way one can speak of an objective fact of the matter as to what a given pocket calculator and a laptop compute, but this objective fact will have as a constitutive feature the interpretative standards and intentions imposed from without by the designers and programmers.[199] When the intentions of the designer become common and shared knowledge, part of the background conditions for interpreting the artificial systems, it is simply taken as a given that the systems do what we take them to do. For practical

[199] This is all well and good for artificial physical mechanisms. But what about natural and organic mechanisms? The formal indeterminacy in such cases should be more obvious and clearer, and if we rid of any 'designer' talk (as we should in the case of natural and organic phenomena) we would still need 'observer' talk since the physical phenomena will satisfy incompossibles. Maybe an appeal to nature or evolution as determining what functions our brain and nervous system was intended will do the trick. Unfortunately, since the evolutionary process is a wholly material process, it will satisfy incompossible 'intended abstract functions or computations' and will be equally subject to the fallacy of sophomoric relativism (see §8). Again, pure functions are not biological functions and the latter will satisfy incompossible cases of the former. The complex processes of evolution can, of course, determine biological function, but not pure (mathematical or logical) functions not the capacity for computing them. The teleosemantic approach (see e.g. Macdonald and Papineau, 2006; cf. Neander, 2018) to human computation, if wholly evolutionary, will be equally subject to the indeterminacy argument.

purposes, indeterminacy issues are of no significant consequence. It is entirely different if the purposes are philosophical and explanatory.[200]

So how do we account for physical computation? And how does it differ from human computation? This is a difficult task. A partial answer to the second question will be given in the next two chapters on the basis of the observation that, though human formal understanding and activity is practice-bound, it's formal (and semantic) determinacy cannot be extrinsic to it but must be intrinsic and inherent to the very capacities, abilities, and acts that make it the case that any competent individual can determinately think mathematically and logically. Human computation is not physical computation. So, the next two chapters will argue for realism about human computation as opposed to non-realism or instrumentalism about human computation. In a way, we save the cogitative powers of Robinson Crusoe: it is possible for him to do mathematics and reason logically in isolation so long as he as the requisite abilities. An answer to the former question, however, requires a work of its own. Partly because any account of physical computation must take into account the varieties of physical computation, which are many. The concept of 'computation' is not the unique property of the cognitive sciences. It is used in fields (e.g., thermodynamics) which have nothing or little to do with the explanation of the cogitative powers of human beings. Our Kripkean argument was directed only at the concept of physical computation in the sense relevant for the explanation of human computation or human formal understanding, although what we say about formal (mathematical) operations being determinate in a way no physical operation can be might be generalizable to the physical as such, with interesting but perplexing consequences for the applicability of formal structures (e.g. mathematics) to matter.

Be that as it may, out of all the philosophical accounts of physical computation I've encountered thus far, the mechanistic account seems to me to be the most promising (Gillett, 2007; see e.g. Milkowski, 2013; Piccinini, 2017, 2015) though it will not deliver the sort of objectivity (realism) their proponents hope for. Take, for example, Gualtiero Piccinini's schema for a mechanistic account of physical computation,

> **The Mechanistic Account of Computation**. A physical computing system is a mechanism whose teleological function is performing a physical computation. A physical computation is the manipulation (by a functional mechanism) of a medium-independent vehicle according to a rule. A medium-independent vehicle is a physical variable defined solely in terms of its degrees of freedom (e.g., whether its value is 1 or 0 during a given time interval), as opposed to its specific physical composition (e.g., whether it's a voltage and what voltage values correspond to 1 or 0 during a given time interval). A rule is a mapping from inputs and/or internal states to internal states and outputs.
>
> (Piccinini, 2015, p. 10)

Piccinini's account is a generic account of computation and is supposed to cover the application of computation or computational explanations over a variety of different fields, many of which have nothing to do with the study of cognition (e.g., thermodynamics).[201] The account is very rich and we

---

[200] "Usually, when a machine is said to compute certain functions, which functions it is said to compute will depend on the intentions of its designer and of its users. Usually these intentions are fairly well fixed, and although in practice the machine may deviate from these intentions in certain systematic ways, such a deviation arises so seldom in practice that we do not worry about it as being the right way to consider the machine. But it is really only relative to an abstract program that is in the mind of the designer of the machine that a machine means anything at all by any of its symbols: addition, multiplication, or whatever." (Kripke, 1984, p. 9)

[201] Piccinini makes an important distinction between explanatory and non-explanatory computational descriptions, the former corresponding to computational modeling and computational explanation. "A mechanistic explanation is a

have neither the time nor the space to cover it completely and in a way that will do it justice. But notice that, given Kripke's argument, the prospects of explaining human computation in terms of Piccinini's notion of physical computation will depend on how the concepts of 'teleological function' and 'manipulation … according to a rule' are cashed out. In the case of formal understanding and activity, that is, 'human computation', the teleological function or end of the activity (e.g., applying mathematical induction to prove something about all the natural numbers) and the rules by which the 'physical vehicles' (e.g., numerals, function symbols, etc.) are manipulated will be *formal* (mathematical and logical), in which case if human computation satisfies the mechanistic account in won't be wholly or completely in virtue of some *physical* functional component, and the physically determined structural relations between the functional components, of the biological (e.g., brain) mechanism of the human person that is a necessary condition for human computation. There cannot be an account of physical computation that can be complete in its explanation of human computation, mechanistic or otherwise. This much we can infer from Kripke's argument.

We have argued that for any arbitrarily selected physical system or mechanism there will always be an indefinite number of incompossible formal rules that it will be intrinsically compatible (or incompatible) with *even if the system is not compatible with any arbitrarily selected function or formal rule, i.e. even if the system does not compute trivially or vacuously any and every function*. So we do not hold that a given physical computing system can be said to compute any and every function. Nor do we think that any and every physical thing computes at least one function. Our argument does not entail any such thing. On the contrary, it is much stronger: the argument entails that no physical computing system can determinately compute a definite and unique abstract function *independently of human intentions*. However, this does not mean that there are no physical computing machines, or that since the formal determinacy of physical computation is ultimately observer-dependent that therefore an arbitrarily selected physical thing can count as a physical computing mechanism or as implementing an arbitrary mathematical function. I take the mechanistic account of physical computation to be partly correct: physical objects without the proper physical functional organization and components cannot count as computational. I also agree that the only way to make sense of the concept of physical computation is in terms of the manipulation of vehicles which are wholly physical, but which can but need not (all depending on the intended interpretation and purposes) stand for formal concepts, rules, or what have you. Physical computing mechanisms, on account of their physical functional organization, can rule out as inadequate a host of computational descriptions. And most physical objects, if we follow the mechanistic account, simply do not satisfy the conditions for being a computational mechanism. So not every physical thing can count as computational, *contra* pancomputationalists (e.g. Wolfram, 2002).[202] But our point has to do with those physical systems that can and do count as computational: even if a physical computing mechanism does not satisfy or

---

description according to which a mechanism (e.g., the human body) has certain components (e.g., the heart), the components have certain functions (e.g., pumping blood) and are organized together (e.g., the heart is connected to the arteries in such and such a way), and the mechanism exhibits its capacities (e.g., blood circulation) because it is constituted by the relevant components, their functions, and their organization." (Piccinini, 2008b, p. 206) That mechanistic explanation would be computational if it attributed, or accounted for, the functions of the mechanism in terms of computations. So computational descriptions are explanatory when the phenomena that is described and whose behavior is computationally modeled itself involves computations—thus, computations are taken to be a real feature of the mechanism that is being computationally described and whose computations figure in the computational description. Computational modeling, on the other hand, models the behavior of a phenomenon without attributing any computations to that which it is modeling computationally (e.g., using computers to predict the weather). Thus, computations do not figure on and are not real aspects of the phenomenon that is being modeled; they are only part of the model. For a more detailed exposition, see (Piccinini, 2015, pp. 60–72, 94).

[202] Again, this is not the place to argue in detail for a mechanistic account of physical computation.

implement arbitrary computational descriptions, there will always be a level of indeterminacy with respect to the *abstract* or *formal* computational descriptions the mechanism is supposed to satisfy. And since mathematics, like logic, is abstract or formal in our sense then doing mathematics and doing logic cannot just be a matter of physical computation. We do grant that there is a sense in which a physical computing mechanism can count at a given moment as adding, for example; our point is, rather, that whenever it counts as adding, though it will not count as also computing every other mathematical function, it will still count as computing *some* other incompossible mathematical function: φ's being in a physical state s_i that can count as a state of *adding* means that φ is in a physical state, namely s_i, that can also count as a state of *quadding*, and a state of *skadding*, and so on. The intended design of the mechanism is what can rule out formal indeterminacy, though the physical mechanism *qua* physical remains indeterminate in form. If the physical mechanism is to count as computing a definite function, the indeterminacy gap must be resolved. But our argument shows that it cannot be *physically* resolved. *This* indeterminacy gap is bridged only by means of the designer's intentions—e.g., by looking at the program, rules, etc., it was *intended* to follow. In other words, the gap is *intentionally* resolved. This is what we mean by saying that the abstract function a physical computing mechanism computes is *extrinsically determinate*, not intrinsically so.[203]

To conclude, we side with the mechanistic account in saying that it is an objective fact of the matter that not every physical thing can count as a physical computing mechanism (*contra* pancomputationalism) and that a physical computing mechanism cannot count as computing any and/or every function (*contra* unlimited pancomputationalism); but we depart from the mechanistic account in recognizing that *which* abstract function(s) a physical computing system definitely implements is, ultimately, a matter of extrinsic determination and denomination. Physically ruling out any arbitrarily selected formal descriptions as true does not suffice to determine physically a specific, unique, and definite formal description. Computationalism as a project that aims in part to model or simulate (not duplicate) human cognition (and, in particular, formal understanding) through the concrete (physical) implementation of abstract formal structures is not threatened in any way by our argument. Likewise, our argument does not infringe upon the foundations of computer science, since it does not imply that there are no objective computational powers in different physical mechanisms and no objective fact of the matter as to which systems are more computationally powerful. For we grant that a physical configuration or structure of the right sort is necessary for something physical to be said to compute. But it is not sufficient. Neither computer science nor engineering need be bothered by issues of formal determinacy (at the level in which we've considered this notion) because for practical purposes it does not matter. It matters for philosophical purposes and for acquiring a proper understanding of what 'computation' is in the case of human computation and what it means when it is implemented by a physical mechanism.

---

[203] "Another objection to limited pancomputationalism begins with the observation that any moderately complex system satisfies indefinitely many objective computational descriptions (Piccinini 2010). This may be seen by considering computational modeling. A computational model of a system may be pitched at different levels of granularity. For example, consider cellular automata models of the dynamics of a galaxy or a brain. The dynamics of a galaxy or a brain may be described using an indefinite number of cellular automata — using different state transition rules, different time steps, or cells that represent spatial regions of different sizes. Furthermore, an indefinite number of formalisms different from cellular automata, such as Turing machines, can be used to compute the same functions computed by cellular automata. It appears that limited pancomputationalists are committed to the galaxy or the brain performing all these computations at once. But that does not appear to be the sense in which computers (or brains) perform computations." (Piccinini 2017) This is a slightly different version of the formal indeterminacy of the physical. Take for example a cellular automata model *M* of the dynamics of a brain, call it *v*. Then the problem is not merely that there will an indefinite number of cellular automata models *N, S, O,* etc., that describe differently but equally well *v*, but rather that even if we narrow the range of plausible computational descriptions of *v* to *M,* there will be at least one other model *M´* (that is, qu-*M*) not compossible with *M* which *v* will inevitably satisfy.

Thus, computationalism as a philosophy of the mind *is* undermined by our argument. For the computational theory of cognition purports to *explain*, metaphysically, formal understanding as the physical implementation of mathematical and logical functions, and if it were true, understanding in human beings would be formally determinate only by *extrinsic* determination and denomination—which is to say, that our formal understanding would be observer-relative and interpretation-dependent (dependent on whom?), not intrinsic and observer independent. There could be no such thing as formal reasoning in any intrinsic and underived way. (We will see in the next chapter why the formal indeterminacy of understanding is impossible.)

8. CONCLUSION: THE FALLACY OF SOPHOMORIC RELATIVISM

We have thus far argued for the claim that no physical phenomena can realize, implement, or apply in a determinate way definite formal rules and functions. Since no physical property, process, or mechanism can implement a determinate formal rule, it cannot implement any formal rule—at least on account of its physical configuration alone. The physical configuration of a computing mechanism is a necessary but not a sufficient condition for formal determinacy. This is true both of artificial as well as natural physical phenomena. We know formal structures are not physical structures. But Kripke's insight lies in recognizing that the latter cannot objectively realize or implement in a determinate and observer-independent way the former. There is implicit in any true computational statement about a physical mechanism a relativity to the intentions of its designer and/or of an observer. Therefore, functionalism and computationalism cannot on the basis of the physical properties and functions of a physical mechanism alone make an inference, except fallaciously, to the non-relativity of any computational statement about any such physical mechanism. This is what Kripke called the "fallacy of sophomoric relativism" (see Buechner, 2011, pp. 358–362).

Consider vague statements like 'John is fat' or 'John is tall'. Statements such as these are "on the surface incomplete" because they omit any explicit reference to the "comparison class" that would eliminate the vagueness. Thus, the statement 'John is tall' is intelligible only in comparison, that is, relative to, a standard of 'fatness/thinness'. Thus, we could bring to the surface the comparison class and thus the 'complete' grammatical structure of the statement as: 'Joh is tall relative to height of population *x*'. This would be an objectively true (or false) fact about John, but entirely dependent on and relative to the selected comparative class. John might be considered short according to the standard of 'tallness' of another population. Kripke takes computational statements about the physical to work the same way. Thus, 'incomplete statements', as we are using this expression, refers to statements that include a relativity predicate but which on the surface neither make the relativity explicit nor eliminate it. Here's the superficial predicate schema of a relativity predicate,

$$\Psi(x_1, x_2, x_3,\ldots,x_n)$$

where 'Ψ' ranges over any relativity predicate, and the values of the x variables are whatever the predicate in question is predicable of. Now, according to Kripke, this is the superficial structure of a relativity statement, which is incomplete insofar as it omits that to which it is relatively true or false. The comparison class needs to be specified in order to the statement to be on the surface complete. Therefore,

$$\Psi(x_1, x_2, x_3,\ldots,x_n, y_1, y_2, y_3,\ldots,y_m)$$

where the values of the *y* variables are the so-called "hidden places", that is, the properties of the comparison class which determine the values that the *x* variables can take. Thus, the "values of the *x* variables depend on the values of the *y* variables." (Buechner, 2011, p. 359) Once the hidden places have been filled in, the statement is complete and "no longer relative to anything at all."[204] All the hidden places must be specified to derelativize the given statement. If it is *in principle* impossible to specify all the hidden places and if the specification of one of the hidden places requires the specification of still other hidden places *ad infinitum* then the predicate, and any statement containing it, will remain relativized to something else and will end up in an infinite regress. The values of the original *x* variables will be indeterminate and thus unspecifiable. This, Kripke claims, is what computationalism and functionalism[205] are subject to, namely the fallacy of "sophomoric relativism". (Kripke, 1984, pp. 11-15)

The fallacy of sophomoric relativism arises from taking a relativity predicate and making use (indeed, inferential use) of it as if it were an objective predicate. The idea is that statements that attribute a computation of a mathematical function to a physical mechanism of the relevant sort are relativity statements with hidden places whose complete specification include amongst other things the intentions of a designer or the interpretation of an observer. Consider the claim that Γ computes some function *f*,

$$\text{COMPUTE}(\Gamma, f)$$[206]

Buechner would formulate the form of the computational predicate as,

$$\text{COMPUTE}(\Gamma, f, \text{PG}, R_1, R_2, \ldots, R_n)$$

This says of Γ that it computes some function *f* according to some program PG and the "correspondence rules" R_1, R_2,…, R_n it uses to compute *f* (Buechner, 2011, p. 360).[207] As Kripke's

---

[204] This is misleading. The predicate (e.g. '*x* is tall') is still a relativity predicate, but now one whose full content has been brought to the surface. Context usually takes care of this. But at any rate, the predicate '…is tall' is still a relativity predicate. The point seems to be, rather, that the buck stops with the specification of the *y* variables. Once we specify that to which the predicate is relative (i.e., the comparative class that accompanies the predicate), we have specified fully or completely the content of the predicate and therefore gotten rid of any unspecified relativity. Relativity stops at the specification of the *y* variables. Thus, if John is 6 feet tall and relative to the Mexican population this satisfies the condition for 'being tall', then 'John is tall' is objectively true but only relative to the general Mexican population. In order to assess whether 'John is tall' is true or false the "hidden places", that is, the comparative class, must be specified. Once it is specified, we have in one sense "derelativized" the statement. The comparative class is no longer hidden.

[205] Again, we are talking about computationalism as a *metaphysical* thesis of the human mind, in particular of formal understanding. In other words, we are talking about the claim that formal understanding and activity is a species of concrete physical computation, of the purely physical implementation of formal rules like mathematical algorithms.

[206] Notice that this predicate, as well as the discussion that follows it, can be framed just as well in terms of the predicate "REALIZE(x_P, *f*)" where '*x*' stands for any physical property, process, or mechanism and '*f*' for any abstract functional role or function which *x* is said to realize. Hence, other versions of functionalism fall within the scope of the argument.

[207] I am following Buechner here, but this is admittedly a strange way to put it. Some distinctions seem in order. We have already distinguished between a function and an algorithm for computing that function. There can be many different algorithms for computing the very same function. Now, what is the difference between an algorithm and a program? After all, isn't program is precisely a set of instructions for the performance of a given (set of) operation(s)? Yes. But they are not quite the same. While each function can be implemented using different algorithms, each algorithm can be implement using different programs. It seems Buechner takes the "correspondence rules" to be the rules constituting an algorithm for the computation of the relevant function according to some program (e.g. some formal language). Finally, not all accounts of computational characterize physical computation in terms of the implementation of a program. Again, we do not presuppose this. Our schema is just one specific illustration of a more general schema. The place

argument shows, this is a relativity predicate because what a physical computing mechanism computes is, as a sufficient condition, relative to something other than itself (e.g. the intentions of its designer, the agreed upon interpretation of some observer(s), etc.). Since a designer D_1 will take Γ to compute the identity function and a designer D_2 will take Γ to compute the quidentity or two-counter function in order for the computational statement to be truth evaluable, the intended design (i.e. the intentions of the designer(s)) need to be specified. Otherwise it remains indeterminate or ambiguous. A complete specification of the computational statement would look like this,

$$\text{COMPUTE}(\Gamma, f, \text{PG}, R_1, R_2, \ldots, R_n, \text{D})$$

where 'D' ranges over any relevant designer (the letter 'O' could work just as well for 'observers'). Depending on the designer, Γ will be said to compute the identity function or the quidentity function, to implement an id-program or a quid-program and follow id-conducive rules or quid-conducive rules. Of course, a more accurate and complete specification can be given for what it is for a physical computing mechanism to implement a given function or program. We do not hold that our schema is exhaustive. A mechanistic account of computation can specify the computational statement with all the relevant properties it takes to be both necessary and sufficient for physical computation. Our point is merely that a complete specification or a derelativized computational claim requires, in the case of the physical, the specification of the intentions of a designer or observer, for the relevant verbs (e.g., '*x* computes') and predicates (e.g., '*x* is a computation') are in physically relative and observer-dependent rather than non-relative and observer-independent. The physical configuration and properties of the mechanism alone do not suffice. As has already been shown, this is true of any physical computing mechanism. Substitute 'Γ' for any other physical computing mechanism or physical realizer of abstract structures and the outcome will be, in substance, the same.

So functionalism succumbs to the fallacy of sophomoric relativism. But consider now 'D'. Is 'D' itself a physical computing mechanism? If 'D' is a human and computationalism is true, then 'D' would be the physical realization or implementation of some formal structure or function. Hence, the computational statement would need to be further specified in order to be derelativized. For, *qua* physical realization of formal structures, 'D' would also require a designer or observer as a sufficient condition for its formal determinacy. But, if that designer is again a physical realization or a physical realizer of formal structures, the same process will have to be done again and again, *ad infinitum*. Hence, we arrive at an infinite regress and no computation would be possible. Unless there is something whose computational (and, in general, formal) determinacy is not *derived* from something extrinsic to it but rather is *underived* and *intrinsic*, there cannot be such a thing as formal determination. We will argue (chapter 4) that this is the case for human beings (i.e. underived formality), so human beings cannot be physical computing mechanisms, nor can formal understanding and activity be the physical realization of formal rules, functions, and operations. At the moment, I merely want to point to the fact that functionalism (and computationalism) as a metaphysical thesis about the nature of formal understanding succumbs to the fallacy of sophomoric relativism and implies an infinite regress insofar as any observer, as well as her understanding of the formally determinate function(s) she takes the physical computing mechanism to compute or realize, will be taken to be a token-physical realization of some abstract function or functional role and thus subject to the very same relativity. If one denies that human psychology or formal understanding in particular is the physical realization of pure, abstract functions (whether considered computationally,

---

occupied by 'PG' in our schema can be substituted for 'FN', where 'FN' range over any abstract functional network, or what have you.

causally, teleologically, etc.) then one will not be subject to the fallacy of sophomoric relativism, but then one has renounced functionalism and computationalism about the nature of our cogitative powers—or one has, at least, renounced the claim that pure abstract functions can be physically realized in any determinate, intrinsic way.

What emerges from our argument up to this point is (1) that whether a physical object and process realizes a pure formal structure is, in short, an observer-dependent fact and (2) in the case of the computation of pure (e.g., mathematical, logical, syntactic) functions, which formal structure a physical computing device realizes is, ultimately, dependent upon the intentions of its designer and the function-mapping (and semantic interpretation) he had in mind when designing the physical device. Two further points:

1. If we explain human logical and mathematical activity in terms of the physical realization or implementation of (formal rules for the computation of) mathematical and logical functions, then whether we reason logically and mathematically will depend, ultimately, on the intentions of our designer (more generally, it will be an observer-dependent fact).
2. But this has the unfortunate and incoherent consequence of making our reasoning logically and mathematically a matter *extrinsic* to the reasoning itself. (see chapter 4)

The problem expressed in 1. is what Kripke called "the fallacy of sophomoric relativism".

Now, Kripke's argument against computational functionalism is generalizable. If one holds to causal-theoretic functionalism, for instance, and employs Ramsey sentences (which are no more than pure formal descriptions) or some other abstract formal description for characterizing psychological predicates and phenomena, then whether a physical process satisfies a Ramey sentence *R* will be relative to the intended Ramsey sentence of an observer or interpreter. For the physical process in itself will satisfy incompossible Ramsey sentences, *R* and *R**. Furthermore, any philosophical account that explicitly or implicitly commits itself to the determinate and intrinsic realization of pure formal structures by the physical will be undermined by Kripke's argument. This is what I call *extrinsic realization*, which is the only sense in which the physical can be said to realize or perform a pure formal operation: a pure function *f* is realized by a physical object φ on account of (i) an appropriate physical configuration and (ii) the intentions or interpretation of an observer such as its designer (if applicable).

What follows from all of this? Two main conclusions can be drawn from our version of Kripke's argument:

(1) Whatever realizes a pure function or pure form must have, and be capable of, *underived* formality (as opposed to *derived* formality; cf. Searle's derived vs. underived intentionality).[208]

---

[208] There is a potential problem for good old-fashioned Platonism about formal structures that we have yet to consider. We have shown that the physical cannot be said to have any formal features in an intrinsic and underived way, but only extrinsically and derivatively. But to possess something derivatively is to possess something on account of something other than itself and in a manner that is wholly accidental to its being (where "wholly accidental" means "attributable to, but not determined by, the thing"). By showing that the physical is formally indeterminate, we show that nothing physical can be a possible subject of formal features in an underived way. But by showing that the physical can be formally determinate by extrinsic denomination, we show that the physical can be a possible subject of derived formality (and in a non-arbitrary way). Now, whatever can have formality only in a derived manner is something that cannot have it in an underived manner. We know the physical satisfies this condition. But what about immaterial subjects, substances, operations, acts, or powers? Assuming we can make sense of such a thing, we certainly have not shown that to be immaterial is to have underived formality. But then again this is not our argument. Our argument consists in finding one power, ability, operation, or activity that has underived formality, and on the basis of that and our Kripkean argument conclude that it cannot be purely physical. Kripke's claim that immaterial substances or powers might be equally subject to his argument might be correct, but is beside the point with respect to *our* argument. If he were claiming instead that *any* immaterial power or operation is subject to his argument, then we have shown (chapter 4) that claim to be false by proving that some power (i.e. a certain kind of intellectual abilities and activity) has underived formality and therefore

(2) That realizing a pure function, i.e. thinking in a pure form, cannot be a matter of extrinsic realization, as good old-fashioned mathematical Platonism would have it—otherwise, the realizer would have *derived* and thus *extrinsic* formal features (or, in my jargon, extrinsic "formality")—but rather must be something intrinsic to the realizer.

A question that has yet to be addressed is how all of this has any bearing on the nature of human formal understanding. The next two chapters will attempt to answer that.

---

must be immaterial, that is, not (wholly) physical (we show this by assuming the intrinsic formal indeterminacy of the understanding and proceed to a *reductio ad absurdum*). Furthermore, we can also mount the objection that such a standpoint threatens to deplete the notions of, and the distinction between, derived and underived formality of any intelligible content. For then it is hard to see what it is that (or what it is to be that) from which material and immaterial powers alike derive their formality, since there would be nothing capable of underived formality (unless we assume some form of Platonism; see below). Indeed, it would be hard to see how we could so much as tell what is derived from what is underived, and thus how we could have any understanding at all of the issue at hand. Similarly, there is also the problem that without such a distinction the further distinction between simulating a formal operation and actually carrying out a formal operation makes no sense. Now, if it turns out that there are immaterial substances or powers which can only be capable of derived formality or at least are not capable of underived formality, then those immaterial substances or powers would be *in se* formally indeterminate. The problem for Platonism about formal structures is that it seems to entail that *everything* except that which belongs to the Platonic realm has derived formality. Only the third realm, say, possesses pure forms in an un underived manner and thus underived formality. But then everything else would be formally indeterminate, including our logical and mathematical understanding, insofar as everything else would be capable only of derived formality—be it by exemplification or instantiation of the pure forms in the third realm. Any formality had by anything belonging to a realm other than the Platonic realm would have formality only in virtue of an extrinsic relation to the latter realm. But if this is so, an incompossible Pure Form could always be said to be exemplified instead by whatever is taken to exemplify a Pure Form. Fortunately, this is an issue that lies well beyond the scope of our investigation. But there is little doubt that our argument, if it does not indirectly refute Platonism about formal structures, it at least brings to the fore a tension that Platonists must resolve in order to do justice to, and account for, human formal understanding if such understanding is underived and intrinsic.

4
Ross and The Formal Determinacy of Understanding

In the preceding chapters, we examined the Kripkean argument for the formal indeterminacy of the physical. Anything physical whatsoever is bound to be intrinsically indeterminate with respect to pure functions, and as such cannot be said in any intelligible way to be capable *in se* of reasoning logically and mathematically. For reasoning logically and mathematically consists, in part, of reasoning in ways determinate in form. At most, what the purely physical is capable of is *approximating* such pure forms and rules or *simulating* logical and mathematical operations, but it can never actually *be* (a case) of any such form and it can never actually *do* the activity in question.[209] "Just as rectangular doors can approximate Euclidean rectangularity, so physical change can simulate pure functions but cannot realize them." (Ross, 1992, p. 141) So there cannot be purely physical operations that intrinsically realize determinately any pure form. Consequently, physicalism about (the doers of) operations that realize pure formal structures is untenable. We have seen that computational functionalism is an untenable philosophical account of some of our cogitative powers (i.e. formal understanding and activity). We have also shown that any theory that aims to account for formal (e.g. mathematical) activity in terms of the (token) physical realizations or implementations of formal structures does not satisfy criteria for intelligible explanations of formal activity. Thus, insofar as variants of non-computational functionalism and non-functionalist computationalism are committed to the determinate and mind-independent realization or implementation of the formal by the physical, our Kripkean argument undermines them equally well (*mutatis mutandis*). One can drop the commitment to the physical realization of the formal and claim that human psychology ought to be explained in entirely different terms. This is as it should be. But then there is no longer a physicalist, whether reductive or non-reductive, account of formal activity. One can also drop talk of abstract functions or pure formal structures altogether, but then one would leave unaccounted that part of human psychology that is actually formally determinable (e.g. learning and mastering algebra) and therefore our capacity for understanding and applying pure formal rules and functions.

One of the main difficulties at this point is establishing what exactly follows (and what exactly does not) from this argument. Suppose you accept the Kripkean argument for the formal indeterminacy of the physical. You may now, *à la* Dennett (see Dennett, 1987), argue that since everything is physical and the physical is formally indeterminate, we have no reason to refrain from inferring that human understanding is, and must be, equally formally indeterminate. Human beings are, after all, animals, and as such they are intrinsically material beings. So, if anything physical is incapable of intrinsically carrying out formal operations, then human beings on account of their nature as material beings are intrinsically incapable of carrying out formal operations. Hence, humans cannot realize or reason in accord with pure functions; at most, their intellectual activity can only approximate or simulate such pure functions. Our treatment of the fallacy of sophomoric relativism should make clear at least one way in which such a response will fail: whether they are simulating or not will likewise be relative to an observer, who will in turn be subject to the same

[209] One can plausibly accuse functionalism and computationalism of being unable to make sense of the distinction between *simulating/approximating* (i.e. imperfectly realizing) a formal operation (i.e. behaving *as if* the function were being realized, mere conformity to the rule) on the one hand and actually *carrying out* the formal operation, i.e. realizing the pure function on the other (what Searle called the difference between "simulating" and "duplicating".)

relativity, and so on *ad infinitum*. We will come back to this later. There are, however, further ways of showing the unintelligibility of such an 'instrumentalist' position about the formal determinacy of human understanding and activity

1. KRIPKE'S ARGUMENT AND FOR FORMAL UNDERSTANDING

The first thing we need to consider is what would be true of formal understanding if functionalism and/or computationalism coupled with token physicalism were true. Consider: the application of formal rules on the part of human beings would consist in the token-physical realizations of abstract rules or functions. But since any physical computing mechanism satisfies incompossible formal rules or functions, it follows that the application of formal rules by human beings equally satisfies incompossible rules or functions. Therefore, human activity in regard to formal operations is formally indeterminate. The only way to determine what is being applied or done (if anything) is by engaging in abstraction and idealization (but who does the abstraction and idealization?). But how a given set of physical processes or structures are formally idealized depends on what *we* take that those processes or structures to be carrying out. After all, it is us who idealize and abstract. Therefore, which function a person computes or which formal rule a person applies is dependent on which idealization is true of her. But which idealization is true of her (as with anything physical that is said to carry out some formal operation) depends on what we intend her (or interpret her) to compute or apply. If, furthermore, the person is identified with a physical computing mechanism, then which formal operations she carries out depends, ultimately, on the intentions of her designer. If, however, she is conceived on a non-computational model but still one in which she is a purely physical being who on account of physical properties alone engages in formal activity, then which formal rules she applies, and which functions she computes will be ultimately dependent on and relative to the interpretation of some observer(s). Of course, the observer(s) or designer(s) will be subject to the same analysis if it's a purely physical being, process, or mechanism, and so on *ad infinitum*. As preposterous and absurd as all of this sounds, it is precisely what follows if token (not to say type!) physicalism about formal (e.g. mathematical and logical) activity is true.

Kripke, similar to the way in which Searle objected to Quine's indeterminacy of translation (Searle, 2002, 1992, 1987), countered that this must be taken as a *reductio ad absurdum* of the premises of the argument. An argument that implies that we never and cannot really reason or think in any determinate logical and mathematical way must be mistaken.[210] Given Kripke's argument, we know that if our reasoning logically and mathematically consists in the physical realization of logical and mathematical forms, which are abstract, then our reasoning in such a way is a matter of interpretation (e.g. of the intentions of our designer?), a derived and observer-dependent fact about us. This only follows if one assumes that what applies to physical things applies to human intellectual abilities and activity, i.e. if one is a (e.g. token) physicalist about formal understanding. Kripke, however, was content to leave this at the level of intuitions. It seemed obvious to him that we just do reason in formally determinate ways (or, in his language, that there is an objective and non-observer fact of the matter as to which mathematical function we are carry out when we carry out a given mathematical function.) But he put forward no argument as to why someone could not countenance in an intelligible manner the view that human understanding was formally determinate and observer-relative.

Enter James Francis Ross (Ross, 1992, 2008). Ross aims to bring home two points:

---

[210] If natural languages are constituted, at least in part, by some definite formal rules, then Dennett's retort would imply that linguistic understanding and activity is an observer-dependent fact. Therefore, we do not objectively speak and understand a definite language; only instrumentally so.

A. The formal indeterminacy of the physical does not imply in and of itself that human understanding is likewise formally indeterminate—one needs the additional premise that human formal understanding (and consequent intellectual activity) is wholly physical.
B. But if formal understanding is wholly physical, then it must be formally indeterminate. Yet any argument for the formal indeterminacy of intellectual activity implies that intellectual activity must be formally determinate; hence, human intellectual activity *is* formally determinate. Hence, intellectual activity cannot be (wholly) a physical process.

It is important to note that, in arguing for the formal determinacy of human understanding (or 'human thought, as Ross preferred to call it), Ross does not seek to point to a fact or state of affairs independent of the pure functions themselves that may allow us to settle, definitively, when someone is (or is not) carrying out a formal operation. The Kripkean skeptic (see chapter 2) is not answered on her own terms but is rather led to see that her skeptical argument cannot be correct when applied to human formal activity because there's no such thing as an intelligible formulation of the *formal indeterminacy of human understanding*. The skeptic's argument is instead shown to be misconceived and incoherent, for there cannot be any intelligibility to her proposal. Ross's observations are not epistemological, but rather ontological and conceptual. The question of whether human beings *can* and really *do* think in formally determinate ways is resolved by showing the absurdities to which we are inevitably lead if we deny that we do. By taking the same considerations that apply to the physical as applicable to human formal understanding, the Dennett-like objection for the formal indeterminacy of understanding succumbs to the impossibility of its being intelligibly raised, precisely because in order to raise and formulate such a claim one must (be able to) do what the very claim affirms one cannot do. The result is that the sort of considerations that apply to the purely physical cannot be intelligibly extended and said to apply to formal understanding nor can they be intelligibly extended and applied to the kind of activity that is the exercise of such kind of understanding.

Here's KRA in syllogistic form:

(P1) No physical thing or process can be formally determinate.
(P2) Some human understanding is formally determinate.[211]
(P3) Therefore, some human understanding cannot be (wholly)[212] physical.

As we've already seen, premise (P1) falls within the proper purview of Kripke's argument. It is the aim of Ross's argument to establish premise (P2). Our adaptation of Kripke's argument owes much to Ross in focus and terminology. But now we move to consider his argument. It is, in my opinion,

---

[211] Ross formulates the second premise of the argument in terms of "some human thoughts" or "some human thinking", such that:

(P1) No physical thing or process can be formally determinate.
(P2*) Some human thinking is formally determinate.
(P3*) Therefore, some human thinking cannot be (wholly) physical.

But the sort of thinking he has in mind is the sort of thinking that issues from formal understanding, so his formulation of the argument is fundamentally the same as ours. Additionally, whether the conclusion of the argument is one that can be said to apply to all of human understanding and any intellectual activity is something we do not investigate here. Ross thought it could, but limited himself to proving it only with regards to the pure forms of understanding. We shall do the same.

[212] "But in part, yes, in the sense that my utterances and actions are physical. Moreover, human thought is not possible apart from feeling or sense, just as a gesture is not possible without bodily movement or inhibition. The target here is theories that thoughts are 'no more than' functions determined physically; for, of course, they are at least 'physically mediated' for humans." (Ross, 2008, p. 198 n. 11; cf. Ross, 1992, p. 137)

the easiest part of the KRA argument—the hardest part being establishing the formal indeterminacy of the physical. The conclusions Ross draws from it in favor of (P2) are the subject of this chapter.

2. AN INTUITIVE SKETCH: THE FORMAL DETERMINACY OF UNDERSTANDING

Kripke's argument implies that no matter what the discriminable features of a physical process or object may be, no matter which outputs it produces and by which mechanisms it produces them, there will always be a pair of incompatible functions, equally adequate empirically, that the physical process and mechanism will satisfy. But incompossible functions cannot be equally satisfied, that is, satisfied at the same time and in the same respect by anything. Therefore, nothing physical satisfies (i.e. realizes, implements, follows) any definite pure function. But if the human intellect is purely physical, then understanding is a wholly physical process or state. But then human understanding could not be determinate in forms, in which case there would be no formal determinacy in human understanding. Intellectual abilities and activities would therefore be formally indeterminate. Consequently, humans cannot understand formal concepts and therefore think in formally determinate ways. Note that this implies that we cannot grasp formal rules, functions, and operations. There would be no such thing as following formal rules and carrying out formal operations. This is surely absurd. We *do* grasp pure functions, we *do* follow formal rules and carry out the corresponding formal operations. Otherwise there would be no such thing as doing mathematics and logic. Therefore, human understanding *is* formally determinate and thus not (wholly) physical.[213] (We give the full argument below, §4.) Ross's original, condensed statement of the argument is as follows:

> Some thinking (judgment) is determinate in a way no physical process can be. Consequently, such thinking cannot be (wholly) a physical process. If all thinking, all judgment, is determinate in that way, no physical process can be (the whole of) any judgment at all. Furthermore, "functions" *among* physical states cannot be determinate enough to be such judgments, either. Hence, some judgments can be neither wholly physical processes nor wholly functions among physical processes.
>
> Certain thinking, in a single case, is of a definite abstract form (e.g. $n \times n = n^2$), and not indeterminate among incompossible forms… No physical process can be that definite in its form in a single case. Adding cases even to infinity, unless they are all the possible cases, will not exclude incompossible forms. But supplying all possible cases of any pure function is impossible. So, no physical process can exclude incompossible functions from being equally well (or badly) satisfied… Thus, no physical process can *be* a case of such thinking. The same holds for functions among physical states…
>
> (Ross, 1992, p. 137)

[213] "The *trait* that determines the tail of the comet, the trait that 'settles every relevant case, including countercases,' marks the contrast with any physical process: a physical process has no feature that can do that. That grounds my main argument: that a necessary consequence of even a single case of such thinking is something that is logically impossible to be a consequence of *any* physical process, or function among physical processes, whatever. Thus, the activity of such thinking cannot be a physical process, and the ability for such thinking cannot be a physical capacity." (Ross, 1992, p. 140)

This passage must be unpacked. First, here we have both an attack on theories that purport to identify formal judgmental understanding with physical processes (or with types of physical processes) and an attack on theories that purport to account for such understanding in terms of functions among physical states and, by extension, theories which presuppose that physical processes can implement or realize determinately any definite abstract form. Similarly, Ross's argument attacks not merely supervenience accounts, which are subject to Kripke's argument anyways, but also emergentist accounts of formal understanding. Secondly, by 'thinking' Ross means "judgmental understanding—what Aristotle though to be the actuality of the intellect…" (Ross, 1992, p. 136 n. 2) or, rather, "judgmental understanding capable of truth or falsity—what Aristotle thought was related to the ability to understand the way one's actual singing is related to one's ability to do it." (Ross, 2008, p. 197 n. 5) This last characterization is closer to our use of 'understanding'. Notice, however, that formal understanding seems necessary for one to be able to make competent and intelligible judgments about formal concepts, rules, and functions. This is no surprise. If one is to judge (competently) whether an argument is a case of *modus ponens* or whether *modus ponens* is a deductively valid form of reasoning, one must *understand* the inferential rule of *modus ponens* and what it is to reason in accord with it or what makes a given logical form deductively valid. Judgments about validity are consequent upon understanding validity and therefore partake in the formal determinacy of the relevant formal concepts. Thirdly, the passage also contains the arguments from incoherent totalities and from the impossibility and futility of an actual infinity which we surveyed in our objection to Stabler (see chapters 2 and 3). Briefly put, supplying all possible cases of a pure function is impossible because

(1) the domain and range of a pure function (or the admissible variable components of a structure) are infinite, and no physical mechanism can produce the relevant infinite totality of possible values,
(2) no physical mechanism can produce an actual infinity of values and even if it could, it could not do both what it actually does and what it would've done instead; so potential and counterfactual cases are always possible (to suppose otherwise is to commit oneself to the incoherent claim that something does both what it *actually* does and what it *would've* or *could've* done instead), and therefore
(3) an actual infinity of cases of a given function cannot exhaust all the possible cases of that function and cannot consequently exclude incompossible functions (e.g. contradictory functional counterfactuals); therefore, an actual infinity of computations cannot determine a definite pure form.

The question that suggests itself at this point is why isn't this true of human beings and their formal activity? In a way it is, since no human being can carry out an actual infinity of anything (though potentially it seems we can, at least cogitatively). Let us first consider what would have to be the case if human beings were capable of formal understanding and engaging in formally determinate activity. Consider again formal operations, that is, the application or computation of formal rules or functions. Suppose you are adding at some time *t*. Ross takes 'being truth preserving for all relevant cases' to be a feature of your adding at *t*, that is, your application or computation of the addition function at *t* has the feature of being containing logically and virtually any possible application of the addition function.[214] The function as the form of the activity must be fully present in the single act, with whatever it logically contains or encompasses, otherwise there would be no

[214] By 'application' I mean correct application. The qualification is unnecessary. And incorrect application is simply a misapplication of the rule.

fact of the matter as to which formal act you are in engaged in or whether you are engaged in anything formal at all. You do not need to compute all the possible values of the addition function (an impossibility) in order to add. Otherwise there would be no such thing as adding. The single case, if it's a genuine application, must suffice. For, a real case of pure function is constituted by the relevant invariable component of that function and is therefore determinate with respect to the features of that 'flow' from that component of the act.

Of course, the application of an infinite function like addition does not make the application actually infinite[215] and, likewise, one could say that the application of an immaterial, abstract form does not make the application immaterial and abstract. Fine. Suppose we concede that. But consider: the application of a determinate function must make the application functionally determinate, otherwise it is hard to see how a definite function is being applied or implemented or computed. This is true of the application of any formal rule. The application of a determinate formal rule is a formally determinate application, and it has to be. To suggest otherwise is to suggest that there can be formally indeterminate applications of determinate formal rules (an 'indeterminate rule' is no rule at all). But which formal rule? If the application of a formal rule is indeterminate with respect to the rule, then it is not determinate that a given formal rule is being applied and therefore it isn't being applied at all. Indeed, it is hard to see how it can be said to be an application at all. An application of what? 'An indeterminate application of a formal rule' is precisely the absence or lack thereof of an application of the formal rule. The application must be determinate whether in mathematical or logical form in order for it to be genuinely (rather than metaphorically) mathematical or logical.[216] Otherwise there cannot be, objectively speaking, mathematical and logical operations and activity. When one applies the arithmetical rule of addition or the logical rule of conjunction, one is adding or conjoining. These are things we do, acts or activities that are determinate in structure or form. Now, if the application of a determinate formal rule must be formally determinate in order for it to be an application of the relevant formal rule, and if no physical phenomenon can on account of its physical properties alone be formally determinate, then (the capacity and ability for) the application of a formal rule cannot be entirely physical. This is the point. Ross's argument merely strengthens on the obvious fact human beings do apply formal rules and therefore engage in formally determinate intellectual activity. From which it follows that our (capacity and ability for the) applications of formal rules cannot be entirely physical.

As Ross writes, all of this points to the fact that certain truth-carrying thoughts (e.g. solving a quadratic equation or reasoning syllogistically) do not seem to be explicable as "reductively [and non-reductively] physical," for they "require features no physical thing or process can have at all". (Ross, 2008, p. 116)[217]

> We have no doubt that the operations of a mechanical adding machine and of a personal computer are entirely physical. Addition cannot be type identical with either of those

---

[215] It makes it potentially infinite, but not actually infinite.

[216] This is to be expected, because formal rules as opposed to other kinds of rules, are rules for structuring one's intellectual activity is certain determinate ways for certain determinate purposes like counting, measuring, deducing, and so forth.

[217] "There is a larger and bolder project of epistemology naturalized, namely, that of explaining human thought in terms available to physical science, particularly the aspects of thought that carry truth values, and have formal features, like validity or mathematical form. That project seems to have hit a stone wall, a difficulty so grave that philosophers dismiss the underlying argument or adopt a cavalier certainty that our judgments only simulate certain pure forms and never real cases of, e.g., conjunction, *modus ponens*, adding, or genuine validity. The difficulty is that, in principle, such truth-carrying thoughts cannot be wholly physical (though they might have a physical medium), because they have features that no physical thing or process can have at all." (Ross, 1992, p. 136)

> physical processes because then it could not be done by the other sort of machine. Suppose that addition is identical with a common function among those processes, then the processes would have to realize that function to the exclusion of every incompossible function. But they cannot do that, as the examples above show. So, the machines cannot really (and univocally with us) add. They cannot do what we do. That is the heart of the matter.[218]
>
> (Ross, 2008, p. 120)

Now, those who take the brain (or, rather, the nervous system as a whole) to be the source and locus of formal operations yet do not take it to be a computer will no doubt find Ross's conclusion wanting. But the argument is intended to apply to any material phenomena whatsoever whether natural or artefactual and whether conceived of in terms of processes, or properties, or mechanisms, or what have you. It therefore cuts through the natural *vis-à-vis* artefactual distinction.[219] (We will deal with the reduction of everything to artefacts when we come to Dennett's objection, §3) In fact, it matters little to the argument whether the material entity in question is an artefact (e.g., a laptop) or natural thing (e.g., a human brain), so in this respect our notion of underived, intrinsic formality differs in scope from Searle's notion of underived, intrinsic intentionality. On account of their material features alone, natural entities and processes satisfy incompossible pure formal structures just as much as mere artefacts and artefactual processes, and therefore neither can constitute on account of their materiality alone determinate instances of any given pure formal structure. As the preceding passage shows, one key feature is exclusion:[220]

> **Exclusion**. A pure function excludes any other incompossible pure function from being realized at the same time and in the same respect ("in the same sense").[221]

---

[218] Note that the inverse also applies: we cannot do or at least do not what they do, at least in terms of speed, efficiency, and simultaneous operations.

[219] At one point, Ross seems to be giving an explanation as to why it is the case that matter, at whichever level, satisfies incompossible pure forms and is therefore formally indeterminate. "In simplest terms, that is because physical phenomena are never under a single quantitative relationship." (Ross, 2008, p. 119) It's hard to know exactly what Ross meant by this.

[220] I call this a principle for want of a better term. Ross does not formulate it so. But it can be extracted from many different passages. For example: "…there are no physical features by which an adding machine, whether it is an old mechanical 'gear' machine or a hand calculator or a full computer, can exclude its satisfying a function incompatible with addition, say, quaddition…" (Ross, 1992, p. 141)

[221] When we say that what realizes a pure form must be determinate with respect to and "logically contain" all the possible cases, thus excluding incompossible forms from being equally well realized, we do not mean that whoever applies a pure function cannot ever misunderstand it in certain domains of application or forget how to apply it. Understanding a formal rule, for example, does not require absolutely grasping *all* the possible values or *all* the possible correct cases of that rule, thus excluding all possible misunderstandings. That would be impossible. But formal rules are different from the rules for the use of words in natural languages, because certain formal rules as rules of a calculi can be made such as to exclude vagueness and determine the correctness of all or any possible case whatsoever. "Exclusion of vagueness, however, does not amount to determinacy of sense. The latter requires the exclusion of the very possibility of vagueness. That is incoherent. Like a signpost, a rule (including a rule for the use of a word) is in order if it fulfills its function satisfactorily. A rule can always be supplemented by another rule to prevent actual misunderstandings. But the thought that we can lay down rules that will exclude all *possible* misunderstandings is incoherent. For there is no such totality." (Hacker, 2005, p. 11) This might be true of words, but it is hardly true of *modus tollens* or De Morgan's theorems in first-order propositional logic. One thing is to lay down rules that are determinate with respect to all or any possible cases, another thing is to lay down a rule that would make impossible any possible misunderstanding of the rule. The former does not require the latter. But we agree that even in cases where there is formal determinacy, and thus rules that budged for all their possible cases, it does not follow that in order to understand such rules one must understand all possible cases of such a rule though one should be *in principle* able to understand any possible case. Similarly, the formal

The 'principle' of exclusion is but an application of the principle of non-contradiction to pure functions and their applications—and it is already implicitly included in our principle of determinate realization, (DR). *Exclusion* says that any particular case of a pure function (e.g. carrying out some formal operation *f* at *t*), if it is to count as a real application or instance of a definite pure function, must partake in the formal exclusion of that function. Otherwise, no single case could be determinate in form and no pure function could be said to be genuinely and observer-independently realized. If in a given case incompossible forms cannot be excluded, then a unique one cannot be instantiated, and no *specific* form can be realized. There would be no such thing as following a definite formal rule, no such thing as carrying out a definite formal operation, and no such thing as grasping and computing determinately any pure function. Physical processes and properties violate the principle of exclusion. For example, any a purported physical application of a formal rule will be indeterminate with respect to what is potentially and counterfactually true of it, formally speaking. So, going back to the pair of incompossible functions ⟨addition, quaddition⟩, we see that it is indeterminate whether a physical process is of the form that would've given you '5' or of the form that would've *not* given you '5' but 125 for the pair of integers ⟨57, 68⟩. Again, genuine application of, say, the formal rule of disjunctive syllogism, secures exclusion in virtue of being truth-preserving for *all* and *any* possible relevant cases of *that* form. In effect, the form of the single case must 'encompass', i.e. logically contain, *any* relevant potential and counterfactual cases whatsoever.[222] "Definite forms of thought in a single case are dispositive for every relevant case, actual, potential, and counterfactual. The function is the *form* of the thinking by which inputs yield outputs." (Ross, 2008, p. 118) The 'input-output' talk is a rather infelicitous way of putting it, but the fundamental point is clear: no physical property, process, or mechanism can be dispositive for *every* and *any* possible case of a pure function (as our Kripkean argument shows).[223] This is why, in part, physical laws are *not* formal laws. There is no formal exclusion in matter.

---

determinacy of pure functions does not make misunderstandings or misapplications impossible. Nothing we've said suggests that.

[222] This is Ross's formulation. Ross uses "all" and "every" to characterize the determinacy in question, whereas I prefer the use of "any" to indicate the very same idea. The problem with the use of "all" and "every" is that it lends itself to the impression that there is such a thing as the set of *all* the cases of a function, that is to say, a real extension consisting of all and every case of a given function. There are at least two ways in which to interpret this. One is at the mathematical level of the arguments and values of a function, that is, the domain and range of functions. Thus, the domain and the range of the addition function are *actually* infinite. From this it is determined what's potentially and counterfactually true of any given application, realization, or computation of the function. What does this mean? Another interpretation at the level of the application or concrete computation of a function (i.e. the set of all (merely) possible applications or concrete computations of the function that are either yet to be actualized or non-actualizable (mere *possibilia*)). There is a sense in which it seems like this is the case (say, when we define the domain and range of the sine function or of the identity function). But Ross, like Hacker, takes this to be an incoherent totality, much like the set of all English sentences (Ross is also against mere metaphysical possibilities, i.e. possible but never actualized entities, events, etc.). Does *our* understanding of such pure forms logically contain or encompass *any* possible case of the function or form? I'd say formally and virtually yes, but not actually so. In virtue of the form of the act, it is true of the act that it is of the form that would've (or would've not) given you such-and-such had such-and-such been the variable component. But one does not need to know readily or attentively be considering any possible case whenever one applies the form in a given case.

[223] The notion of "determination" or "determinacy" at play here is not one about 'determinate sense or meaning' but one about 'determinate form', as both (1) and (2) make clear. Some authors (Dillard, Feser) interpret Ross's talk about human understanding being formally determinate as talk about the content of formal reasoning being unambiguous or lacking in vagueness. But this is incorrect, and hinges on a conceptual blunder if not downright ignorance about the difference between form and content. It turns Ross's *main* argument into one about the determinacy of the meaning, content, or sense (*Sinn*) of thinking, when it is in fact primarily and clearly about the determinacy of the form of thinking (say, inferences, calculations, judgements, and so forth). The semantic interpretation cannot explain why Ross focuses on

Ross pays special attention to truth-preservation but seems to understand it as a general formal feature (e.g., a feature predicable of mathematical functions) rather than a feature only of deductively valid arguments. In our jargon, anything that is truth-preserving in virtue of form is thus determinate with respect to (i.e. logically contains or encompasses)[224] *any* possible transformation of the form's variable component, thus effectively excluding any other incompossible pure function from being equally well carried out or realized at the same time and in the same respect.[225] It is the invariable component, not the variable component, that we take to be the form of the operation or application (see chapter 1).[226] To 'logically contain any possible case' is for it to be such that it determines what accords with it and what is to act in accord with it for any possible application or computation—say, the way '$n + n = 2n$' logically contains the values for any possible arguments of the right sort (cf. Ross, 2008, p. 92). We say that '$n + n = 2n$' logically contains or encompasses '2 + 2 = 4', '3 + 3 = 6', '10 + 10 = 20', '624 + 624 = 1248', and so on for the infinity of possible cases. No one of its cases is more or less of a case of that form than any other (e.g. '3 + 3 = 6' is just as much as case of '$n + n = 2n$' as is '10 + 10 = 20'). In other words, both cases are cases of the pure function '$n + n = 2n$' and is therefore true of a particular case or instance of '$n + n = 2n$', e.g. '2 + 2 = 4', that it is of the same form that would give or would've given '20' had the value of n been '10'. A particular case (i.e., a particular, concrete application or computation) of a mathematical function must partake in the logical containment of the function. Otherwise, the particular case could be made out to accord with incompossible mathematical functions and there would be no definite function which the particular case could be a determinate computation or application of.

---

form almost entirely, nor why (2) would be an essential element for proving that logical reasoning cannot be materially realized (since the issue is truth-preservation *in virtue of form*, not content). This misrepresentation of Ross's main argument fails to appreciate both the originality and the implications of his argument and makes Ross's argument vulnerable to objections (e.g., the Pasnau objection) that it should not be vulnerable to. See §6. Suppose the order is to follow the rule 'add 2'. Hacker asks, "does giving the order 'Add 2' and meaning it in such-and-such a way consist in performing an infinite number of acts of meaning correlating every even number with the next even number in the series? … An infinite number of acts of meaning is absurd, but why does the resolution of the difficulty not lie in the performance of a *single act of meaning* which will *contain* all the acts that accord with the rule in advance of their performance?" (Hacker, 2009, p. 71)

[224] It doesn't *actually* contain them but only logically, that is, in virtue of its definition. The actualist viewpoint is a Platonist one. It is not necessary.

[225] The question is 'how does it contain them'? In what sense? This is beyond the scope of our present investigation. For a controversial answer to this question, see Hacker (2009). We will address the issue of non-standard models of first-order logic in the next chapter.

[226] One can, of course, decide to keep invariant the content (that is, the variable component) through different formal operations (say, keep ⟨3, 3⟩ invariant through different arithmetical operations, e.g. go from "3 + 3" to "3 × 3") and in this sense one could say that the content of the operations is what remains invariant through transformations of the form. But this is no counterexample to our stipulation of the form as the invariable component of a formal operation, because the example constitutes a change *in what is being done* (that is, a change in the form, in the invariable component), though not a change in what it is being done to, and we are interested in that which doesn't change through suitable transformations of its variable component. To go from one operation to another without a change in content is to substitute one invariable component for another that are not equivalent and therefore it is to go from doing one thing to doing something another over the same thing. If in such cases one wishes to call the arguments or the content (e.g. ⟨3, 3⟩) of an operation "that which can remain invariable through transformations of the form" then alright. But that would be an issue different from the one we are tackling, for what determines that different operations have really taken place? The content is the same. The output? But incompossible functions can yield the very same output on the basis of the same content (i.e. arguments) for a segment (finite or infinite) of their domain. Then what determines it? It is the *form* to which we must look, for it is there that the differentiating principle between incompossibles is found. The output *manifests* the difference, but it does not *constitute* it.

> 'Being truth preserving for all relevant cases' is a feature of the single case of validity. The whole function (e.g. modus ponens) is realized in the single case. It is not like 'being punctual' that is realized over a pattern of occasions. The form of the reasoning that occurs in the particular case is 'truth preserving.' Otherwise it would not be 'impossible by virtue of the form to proceed from truth to falsity' in *that* reasoning. So, the form of the actual 'encompasses' (logically contains) all relevant cases whatever.
>
> (Ross, 2008)

In terms of formal rules, to say that a formal rule logically encompasses all its possible applications is to say that the rule settles normatively what is in accord with it actually, potentially, and counterfactually, and to say that a single application of the rule must logically encompass any possible case is to say that it is not indeterminate with respect to what is actually, potentially, and counterfactually true of it, formally speaking. Recall our characterization of the notion of 'formal structure' (chapter 1) as the component of formal activity that remains invariable throughout transformations or substitution instances of the right type of variable component. A pure formal structure therefore logically encompasses all possible transformations of its variable component on account of that which remains invariable and therefore logically encompasses any possible variation of the relevant type of variable component.[227]

Two points of confusion must be avoided at this point. First, that the form of the thinking is a single case must logically encompass or contain any possible case of that form does not imply that one must have a grasp of any and every possible case whatsoever of that form. This is impossible. I can calculate in a way that is determinate in form, say, when I multiply integers, without any actual knowledge as to what the value of some indefinitely large pair of numbers is or as to the fact that multiplication is commutative for real numbers but not for quaternions. We are not granting a privileged, omniscient access to all that is true of the relevant form of the thinking. Understanding comes in degrees, and so the application of a formal rule will be *with respect to its variable constraint* constrained by the relevant degree of understanding and a host of other factors like memory, attention, etc. Yet the form must be fully present there, structuring the act, and one does not need to know that multiplication is not commutative for quaternions or what multiplication would yield for $\langle 15346^{10x607}, 951^{-25} \rangle$ in order to multiply, say, the pair of integers ⟨-8, 9⟩. But the single case of multiplication over integers or of squaring an integer, for example, must be of the definite abstract form (that is, an instance or application of the form) '$x \times y = z$' or '$n \times n = n^2$', respectively, which

---

[227] Don't natural forms also then satisfy this condition of logically containing all possible cases whatsoever? I'm not sure, but if they do, in won't be in the same sense as pure forms and it won't be logical containment. I'd say natural forms contain them naturally, in potency on account of the natural substances they are and what can be actualized when some of their powers or capacities are engaged. This is material containment the way potential offspring are contained in the genes of members of a species together with a host of other material (e.g. biological) factors. But the offspring is not determinate until it comes about (although Platonists might disagree). In addition, it cannot contain all possible cases actually—otherwise all actually existing material structures would be all the structures that could be. The actual cases of a form do not exhaust the form. (This is true of any form). Logical containment, however, is determinate with respect to any possible case whatsoever on account of the invariable component rather than on account of the variable component or accidental features of the cases (say, a kid in the Dominican Republic adding for the first time at 2:24pm local time on February 6, 2021). It is not clear that what is accidental in a given case of a pure form is accidental to a given case of a natural form. At any rate, though there are similarities and overlaps between all categories of structures, formal structures are normative in a way that is practice-based and intellectually bounded (or bounding) and are can be applied for the study of certain phenomena, even phenomena not within the same ontological category (e.g. material substances and processes). No such thing applies to natural forms. Mixed formal structures are an in between, yet we will not consider them here however interesting.

logically contains all relevant cases whatsoever and settles what is both potentially and counterfactually true of the single case (and, concurrently, what is actually true).[228] Secondly, formal (e.g., logical) containment should not be confused with *temporal* containment, that is, the claim that a single case of a pure function contains all past, present, and future cases of that function. This is conceptually awry, and I'm doubtful that it is intelligible. The single case (say, my adding right now) is a particular concrete act that is partly constituted by a form or structure that is pure, that is, complete and definite independent of any material property or materialization. Pure functions are not spatiotemporally bounded yet their applications are. How could this be? Ross does not really address this issue. But my response would be to deny that the applications of formal rules are *wholly* spatiotemporally bounded and determined, otherwise they would be wholly physical. They are, of course, *partly* bounded by spatiotemporal dimensions insofar as our mathematical and logical activity is embodied (e.g., we learn and apply formal rules while at a given place and at various moments in our lifetime, it takes time to carry out a given formal operation, it can be tiring and difficult to prove a theorem, etc.). But mathematical and logical activity must transcend the limiting bounds of spatiotemporal dimensions, otherwise there could be no such activity. We could say that the spatiotemporal constraints belong to the material, not the formal, cause of our formal operations and activity, but this would require a suitable metaphysical framework (e.g., maybe an Aristotelian causal pluralism of the sort proposed by (Jaworski, 2016) and/or a Wittgensteinian account of rule-following and rule-understanding in mathematics and logic) to flesh this out; but this is beyond the scope of the present investigation (but see the next two chapters and the conclusion for some ideas).

Whether you consider pure formal structures as *eternal* (Platonism, Augustinianism) or as *sempiternal* (Aristotelianism, Wittgensteinianism), or as neither (Fictionalism), what cannot be intelligibly denied is that formal activity, *qua* the exercised ability to follow and act in accord with pure formal rules, requires that the particular relevant acts—which are no doubt embedded in spatiotemporal dimensions—be formally structured, and therefore that the formal structure be wholly present in the particular act and be *intrinsic* to it. But this does not mean that a pure function determines all "its" future or past computations, because a pure function has no temporal determination. That makes no sense. Consider set-theoretic determinacy. Let *N* be the ordered set of all the natural numbers. Then one can talk about every or any possible member of that set, about which numbers are prior to which, about which numbers must be computed first if one is to compute pairs of natural numbers in a given order (say, from smaller to larger), etc. But the order is not temporal, and neither is the containment of the numbers. The larger numbers are not future

[228] There is controversy about whether multiplication over quaternions is univocal with multiplication over integers, and even whether multiplication over positive integers is univocal with multiplication over negative integers. Some argue (Hacker) that the introduction of negative integers into the domain of the multiplication function constituted a change in the meaning of 'multiplication', and change in the formal concept, and the same can be said in the case of quaternions. Others might simply hold that it was a change in the extension, not the intension, of the operation and therefore that multiplication is univocal—except perhaps in the case of quaternions where commutativity does not hold. If it's not univocal, then fine. Our application will be determinate with respect to the relevant sense of the formal operation in question. A new operation will have to be taught for multiplying quaternions. It doesn't matter. Multiplication considered as a function from natural numbers to natural numbers is itself infinite, and we can make the argument with respect to multiplication in this sense. If it's univocal, then fine too. Then it would remain true of a single case of multiplication over natural numbers that it is of the same form that would have given you such-and-such had the *type* of numbers been different. When we teach our logic students the logical rules of first order propositional logic, they might not know *that* nor even *how* these rules are applied in, say, predicate logic or that these rules are constitutive of a formal system that is complete and consistent (i.e., first-order propositional calculus). But they will know what it is to apply them and therefore will be able to apply them to the extent that the introductory course requires students understanding of these rules. But their teacher will know that the rules that they are applying and following have features and applications that the introductory student does not, will not, and need not grasp at the moment. There is no problem here.

cases, nor are the smaller numbers of the set in any temporal sense in the past relative to the larger ones, nor is the notion of 'prior' or 'order' a spatiotemporally (and therefore physical) one. Now consider a non-partial *n*-ary operation or function over *n*-tuples of *N*, e.g., the addition function or the identity function. Then the identity function is determinate with respect to every member of that set, both in its domain and range. This is formal determination. Similarly, in our case, the addition function has a domain and a range that is complete and fully determinate (namely, *N*) such that for any possible pair of arguments taken from *N* the addition function has a unique value (namely, one member of *N*) corresponding to them. There is no spatiotemporal determination here. A certain tectonic movement might determine certain immediate, or less immediate, future events and physical cause others. But there is no sense in which a set of physical properties can determine and cause the (future) values of a function. There would be a categorical conflation here.

Such a conceptual confusion, I think, seems to rest on a misunderstanding of the nature of our applications and computations of pure rules and functions, that is, of the nature of formal operations. Surely, our formal operations and activities are temporally bounded. For if we carry out a given formal operation we carry it out at some moment in time and in certain place, in a finite amount of time, using certain forms of representations and images (whether notational or mental), and so on. If a particular formal act at *t* counts as an instance of adding or of some algorithm for computing the values of the addition function, then this is so because the application or operation at *t* is of a given definite form *f* and therefore has the pure invariable component of *f* as the form of the calculation at *t*. But this is not to say that formal operation *f* at *t* must in some spatiotemporally causal way (say, on account of some property of *t*) determine what is to be done at some other time *t** (where *t** ranges over any moment in time other than *t*) if one is to carry out *f* at *t**. The present application does not *cause* future applications in any way, nor are past or future calculations relevant for the determination of the application of a given formal rule at some moment in time. Otherwise there could be no such thing as applying a given formal rule for the first time (e.g., a logic bachelor student learning De Morgan's theorems and applying them for the very first time). A pure function or rule normatively determines what is in accord with it (and what is not in accord with it) for any possible application or instance of it, and it does so independently of *when* and *where* the rule is to be applied or has been applied. That's why we don't need to consult past or future applications in order to apply a given formal rule. If one understands the rule, one will know what it is to apply it and will have the ability to do so. That's it. Spatiotemporal dimensions and constraints are accidental to the *form* of the application, though not to the matter and means of the application—that's why it does not matter *with respect to its formal and normative nature, to the validity or the soundness* of the proof, whether it took Gödel a year or three months or 6 hours to prove his incompleteness theorems; likewise, as mathematicians and linguists alike have recognized, spatiotemporal constraints are accidental to our recursive capabilities since we could *in principle* carry out a recursive operation without end and we can understand this to be so, even though we cannot actualize it. I take this to be Ross's point when he says that the spatiotemporal features of formal activity are accidental to the *form* of that sort of activity. To quote the passage in full:

> Size and length of a computation … are external to the form of thinking, accidental to what is done … The fact that I cannot process every case of *modus ponens*, because most of them have premises too long for me to remember, sentences too long to say, or words I do not understand, is adventitious, like my not being able to do *modus ponens* in Portuguese. Those are features of the functors, not of the function. The function that has to be realized in every case is the one wholly realized in the single case. The point is to be taken literally: that the function is wholly present, not by approximation, exemplification, or simulation, but by realization in the single case … consider an even

> simpler function, "conjoining." Conjoining is the functional arrangement of an *n*-tuple of assertions into a single assertion that is determinately true just in case every one of the *n*-tuple of judgments [i.e. the conjuncts] is, and false otherwise. The truth of the whole block is the truth of all of the units ("$p \cdot q$ = T just in case p = T and q = T"). I can conjoin every sentence in the fourteenth edition of the Encyclopedia Britannica, or yesterday's *Times*. What I do in the single case is what would conjoin any string of suitable units, even ones too long for me to think of, or beyond my access to refer to. It is impossible to conjoin thoughts, if what I do is indeterminate among incompossible forms (at the same level).[229]
>
> (Ross, 1992, pp. 138–139)

> It [conjoining, for example] is a manner of *understanding*, not just a grammatical form or a logical transaction. It is something we do, like thinking "I had the appetizer and the entrée." This is not about how many states we can be in. It is about the ability exercised in a single case, a definite thought form distinct from every other. It is all there, all at once.
>
> (Ross, 2008, p. 118)

So this is what has to be the case if human beings possess formal understanding and engage in formal activity. Suppose I am now adding. Then what I am now doing is formally determinate in the specified sense and therefore determinate with respect to any possible case (counterfactual or potential) of the formal rule. But this is in no way a temporal, causal kind of determination. This is what I take Wittgenstein to be aiming at in a passage cited and discussed by Kripke, albeit with reservations (Kripke, 1982, p. 70 n. 58):

> *PI* §I89: "But are the steps then *not* determined by the algebraic formula?" In spite of Wittgenstein's interpretation within his own philosophy of the ordinary phrase "the steps are determined by the formula", the impression persists that the interlocutor's characterization of his view is really correct. See §I95: "But I don't mean that what I do now (in grasping a sense) determines the future use causally and as a matter of experience, but that in a queer way, the use itself is in some sense present," which are the words of the interlocutor, and the bland reply, "But of course it is, 'in some sense'! Really the only thing wrong with what you say is the expression "in a queer way". The rest is all right; and the sentence only seems queer when one imagines a different language-game for it from the one in which we actually use it."

The form is present, all there, in the single case, not in a queer way but in a straightforward way (see Ross, 2008, p. 118). Except in the trivial sense of 'if one were to add in the future one would do then the same thing one does when one adds and therefore what would give you such-and-such values given such-and-such pair of numbers', there is no sense in saying that a formal rule or that understanding a formal rule determines what it is to accord with it for all future applications if this is

---

[229] By the qualification "at the same level", Ross refers to the same level of specificity concerning the form of the inference or act. We know that arguments in which predication plays a constitutive role in regard to the validity of an inference turn out invalid in formal systems that do not accommodate predication (e.g. standard propositional logic). One needs a 'richer' formal apparatus to capture the form of the inference at its most particular, or specific, level. Hence, the need for predicate logic. Ross later refers to forms considered at that level as "equally most particular forms." (see e.g. Ross, 2008, p. 197 n.10) The qualification is presupposed throughout.

taken to mean temporal determination rather than the obvious sense that since a formal rule determines what is in accord with it for any possible transformation of its variable component, any application of the rule at any time whatsoever will be normatively constrained by it.

Much of what I've said thus far is more of my own ruminations on the basis of Ross's at times obscure, at times incomplete, comments. Such will be the nature of our Rossian argument. But we supply ample textual evidence in support of the coherence between Ross's argument and my own ruminations. Let us end by taking a fresh look at our principle of the determinate realization:

> *Principle of Determinate Realization (DR)* = For any realizer *P*, *P* realizes the pure formal structure *f* (where *f* is a variable ranging over pure functions) if and only if all the features of *P* relevant for the realization of *f* determine that it realizes *f* and only *f*, such that for every *f´* that *P* realizes, *f´* = *f*.

This principle says that for something to constitute a real or genuine realization of a pure function *f* the properties or features relevant for the realization of *f* must fully determine that *that and only that f* is being realized. This must not be read as saying that simultaneous computations are impossible or that the same set of physical properties relevant for the computation of one function in a given computing device cannot be the relevant for the computation of another function. Our qualification 'at the same time and in the same respect' or 'in the same sense' must not be ignored. In the case of physical computation, our principle should be taken to state a rather plausible requirement: the physical properties of a physical computing mechanism responsible for the computation of a mathematical or logical function (or a set of mathematical and logical functions) must be determinate with respect to which function (or which set of functions) it computes, otherwise it is hard to see how there could be real, determinate computation by the physical of *definite* mathematical or logical functions.[230] This is an intuitive sketch of Ross's argument. The thrust of the argument against the formal indeterminacy of human understanding will be developed when we come to the refutation of Dennett's objection.

3. CONTRA DENNETT: FORMAL UNDERSTANDING AND THE INTENTIONAL STANCE

We have made use of the distinction between derived-extrinsic and underived-intrinsic formality, following Searle's (see e.g. Searle, 1992, pp. 78–82) analogous distinction between derived and underived intentionality. Daniel Dennett is known for having objected to this distinction and for having countenanced the claim that there is no such thing as underived intentionality, only derived intentionality. Let us consider Dennett's objection (Dennett, 1996, cf. 1991, 1987)to Searle's *et al.*

---

[230] "A leading idea of modern mathematics (and much philosophy of science) is that any function is under-determined by a proper subset of pairings of values with arguments. This idea seems immediately applicable to the rule-formulation for writing the series of even integers when it is given in the form '0, 2, 4, 6, 8, 10…' With sufficient ingenuity, some function can be specified which yields as its values any continuation of '0, 2, 4, 6, 8, 10…' for the series of arguments 6, 7, 8, … This seems to imply that writing any sequence of numbers accord with the rule formulated by '0, 2, 4, 6, 8, 10…' From a logical point of view, it seems, the formulations of a rule by examples leaves wide open what accord with the rule. On *some* interpretation of the rule *any* continuation is correct." (Hacker and Baker, 2014, p. 91) This is, in a sense, true but false in another. In the sense in which a formulation is physical or material (e.g. *qua* written), it leaves wide open which rule it accords with. Even if we grant that '0, 2, 4, 6, 8, 10…' written on a sheet of paper stands for a series containing the first five successive even numbers, starting from '0', the written "sequence" leaves it open whether it is a case of '+2' or of some other incompossible function. But a *formulation* like a definition of a formal rule is not a purely physical or material phenomenon, and therefore a formulation of a formal rule if it is a formulation of that rule will not be indeterminate. Again, formulations of pure functions are things we (can) *do*.

notion of underived intentionality and then proceed to apply it to the formal aspects of understanding. Suppose there is a vending machine that, although it can discriminate between multiple types of coins, cannot discriminate between a US quarter and some other kind of coin, call it qu-balber. Is it a quarter-vending machine or a qubitser-vending machine? If what makes the machine a qu-balber or a quarter vending machine is not, as a sufficient condition, any of its intrinsic causal-structural properties (i.e. it's physical architecture, is outputs on the basis of certain inputs, etc.) then what makes it a qu-balber as opposed to a quarter machine or vice versa is extrinsic to it. When what is extrinsically predicate of something is dependent as a sufficient condition for falling under the predicate determinately on human conceptualization, intentions, interpretation, and so forth, that something is said to have that property in a *derived* rather than in an *underived* manner. (By derived 'property' it is meant a property had not on account of something *intrinsic* to what has it but on account of something extrinsic to it which fixes and bestows the possession of the property to the thing that is said (on account of something other than itself) to have it or to lack it.)[231] Thus, what the vending machine or what its behavior and internal structure *represents* (e.g., a qu-balber receptor) is ultimately settled by the intentions of its designers, owners, or what have you. Physically speaking, nothing suffices to make the machine a qu-balber vending machine as opposed to a quarter machine, and vice versa. Context, utility, interests, and interpretations determine which kind of machine it is and what the machine represents, but intrinsically and underivatively (that is, on account of its physical structure alone) nothing does. Physical mechanisms like these (i.e., artifacts) behave *as if* they had intentionality, *as if* they represented and understood what they were doing, but they do not *really* possess intentionality, do not *really* represent and understand anything and do not really possess the semantic properties that we attribute to them except, of course, derivatively (see e.g. Searle, 1992, p. 78ff for a standard statement of the distinctions between intrinsic, derived, and as-if intentionality).

Dennett responds by reducing all things (including us) to artifacts, metaphysically on a par with objects like the vending machine. For Dennett there is no such thing as *underived* intentionality—at least not in *us*. So, the idea that we, when it comes to intentionality, meaning and conceptual content, are categorially different from artifacts is rejected. Whatever intentionality and conceptual content we can be said to have it will only be derived, not intrinsic. Apparently, Dennett takes "Mother Nature—the process of natural selection" (Dennett, 1987, p. 301) to be the one that bestows, or from whom we derive, our intentionality (and *mutatis mutandis* formality). "So our intentionality is derived from the intentionality of our 'selfish' genes! *They* are the Unmeant Meaners, not us!" (Dennett, 1987, p. 298) 'Mother Nature' is, however, either an artefact or a non-artefact. Presumably, Dennett would want to say that it isn't an artifact. Then the buck stops with 'Mother Nature' and therefore "she" (i.e., natural selection) must be the possessor of original/underived intentionality.[232] Yet Dennett seems to point in the opposite direction,

> First, if we are (just) artifacts, then what our innermost thoughts mean—and whether they mean anything at all—is something about which we, the very thinkers of those thoughts, have no special authority. The two-bitser turns into a q-balber without ever

[231] Note that this characterization of what it is to be a derived property or feature might be too broad, and thus include *as-if* properties as well as derived properties. But this can be resolved by considering *as-if* properties as derived properties that are not real, whereas the formal properties attributed and attributable to physical computing mechanism are real but derived properties.

[232] "So if there is to be any original intentionality—original in the sense of being derived from no other, ulterior source––the intentionality of natural selection deserves the honor. What is particularly satisfying about this is that we end the threatened regress of derivation with something of the right metaphysical sort: a *blind* and *unrepresenting* source of our own sightful and insightful powers of representation." (Dennett, 1987, p. 318)

> changing its inner nature; the state that used to mean one thing now means another. The same thing could in principle happen to us, if we are just artifacts, if our own intentionality is thus not original but derived… if we are such artifacts, not only have we no guaranteed privileged access to the deeper facts that fix the meanings of our thoughts, but *there are no such deeper facts*. Sometimes functional interpretation is obvious, that when it is not, when we go to read Mother Nature's mind, there is no text to be interpreted. When "the fact of the matter" about proper function is controversial—when more than one interpretation is well supported—there is no fact of the matter.
>
> (Dennett, 1987, pp. 298–300)

There is much more in Dennett's objection to Searle *et al.*, but what is relevant for our purposes is the observation that there is no such thing as underived intentionality. If we consider the argument in terms of the semantic aspects of understanding, we can say that Dennett's objection amounts to the claim that ours is derived rather than underived and therefore extrinsically determined.[233] Dennett's strategy is actually quite simplistic and straightforward. It consists in accepting that there might be, in certain cases, such a thing as original, intrinsic, or underived intentionality, but it is not found in human understanding. It is, rather, to be found elsewhere: in the evolutionary process of natural selection that assigned certain functions and contents (or formal capacities—which ones?) to our psyche. Therefore, the human psyche is only capable of derived and extrinsic intentionality. Dennett also grants that there are cases in 'Mother Nature' in which there is no fact of the matter as to the function, meaning, or content a type of mental state is supposed to have. In such a case, there is no fact of the matter as to the content of our cognitive states or concepts and indeterminacy will abound. If we formulate Dennett's objection with respect to formal structures, the argument would be something like this. Which formal operation we (can) carry out is a matter extrinsic and derived from Mother Nature. Therefore, there is no intrinsic and underived formal determinacy to our higher intellectual activity. Take conjunction. We do not really conjoin, that is, think in this way in an intrinsic and underived manner but rather we conjoin (if we do) because conjunction is the function assigned by natural selection and found in our genes. The same goes for adding, simplifying radicals, applying *modus ponens*, etc.

> meaning … is, in the case of an organism, independent of our intentions and purposes, [but] is not independent of the intentions and purposes of Mother Nature, and hence is, in the end, just as derived and hence just as subject to indeterminacy of interpretation, as the meaning in our two-bitser.
>
> (Dennett, 1987, p. 305)

---

[233] There is an equivocation in the meaning of the terms 'intrinsic' and 'extrinsic'. One way is to use 'extrinsic' as meaning outside or as dependent on something other. So, when we see a cat, what we see might be said to be extrinsic to the seeing, for the cat is other than the seeing and determines in part the content of the visual experience. The same holds for natural languages and games if they are essentially socio-conventionally determined (contra Chomsky). This notion of 'extrinsic' is not entirely at odds with our position. See (Dretske, 1993) for a characterization of the poles 'intrinsic/extrinsic'. But there is another way in which 'extrinsic' is used, namely to mean that what is extrinsic is not found or it does not belong to something on account of any of its properties and powers but only on account of a relativization to something other which could be otherwise. Thus, to say that 'formal understanding' is extrinsic to us is to say that we do not (and cannot) possess or have the capacities and abilities for formal activity except in the sense of being interpreted as having them or as being said to have them by some (artificial or natural) functional assignment that can always be otherwise and is thus compatible with incompossible pure forms. Is this latter use of the term 'extrinsic' that is at odds with our argument. No doubt formal understanding is extrinsic in the former sense, since it is dependent on certain formal practices and activities, learning and training, conventions, etc. But the capacity is intrinsic to the person, and so are the abilities (the understanding) once acquired.

It does not take much to show that 'Mother Nature' is formally indeterminate through and through. But Dennett would welcome this, it seems, and simply accept that there is no fact of the matter as to which functions we compute or which formal rules we apply. If that is true, Dennett's own argument will be formally indeterminate and therefore cannot be of any definite logical form (in which case it is hard to see how one can talk about an argument at all). Furthermore, it is just silly that in order to know which argument Dennett is putting forth we need to consult his genes or the human genome, and it is not clear how that would help. At any rate, notice that Dennett cannot escape the infinite regress of the fallacy of sophomoric relativism. When it comes to formal structures, Mother Nature will always satisfy incompossible ones since it is a pure material process and therefore will owe any formal determinacy to something extrinsic to it, in which case it will be derived. But from what? The statement 'Mother Nature has intentionality/formality' remains a relativity statement. Bear in mind, also, that if our thinking is extrinsically of a logical form, then it is only relative to something other than the thinking. But this is just to say that thinking does not really have a logical form at all, intrinsically speaking. So, no thinking is ever a real case of a logical or mathematical form; it is only of a logical or mathematical form relative to something else and therefore always compatible with incompossible pure forms. But then there cannot be any such thing as mathematical and logical thinking, not even of the sort that leads to Dennett's conclusions. We provide our full refutation in the next section.

Notice that derived formality is observer-dependent. What kind of being would count as an appropriate 'observer' such that it can bestow upon our understanding the formal determinacy that it does not *really* have? Dennett would want to say 'artificer-dependent', where the artificer is Mother Nature. Fine. But if Mother Nature is wholly material, as it surely is, then it will be subject to formal indeterminacy, in which case it makes no sense to say that there can be underived, intrinsic, or original formality in Mother Nature. Let us not forget what the fallacy of sophomoric relativism teaches us. Suppose we are purely physical mechanisms in the sense that all of our capacities, abilities, and activities are wholly physical. Then in the relevant respects those aspects of the observer (e.g. her intentions, interpretation, understanding, etc.) relevant for the determination of a given formal structure would in themselves be formally indeterminate and could become determinate only relative to (the intentions or interpretation) of another observer, who would in turn be subject to the same analysis if human. And so on *ad infinitum*. Similarly, supposed we are physical computing mechanisms, at least in certain respects (say, in respect to mathematics). Then, as we saw in the previous chapter, what we do mathematically speaking would be relative and dependent on the intentions of our designer. But if the designer itself is a purely physical mechanism or process, then it will be in need of an observer or designer in order for there to be any intelligible predication of formal determinateness, and so on. Here is where both Dennett (Dennett, 1987, pp. 298–300) and the Intelligent Design (ID, for short) proponent meet and fall. The ID proponent, like Dennett, would happily welcome the conclusion that whatever intentionality and formality the understanding exhibits will be extrinsic to it. But, says the ID proponent, if reasoning in any determinate formal way (as capacity, ability, and activity) is dependent upon the intentions of a designer, then if our reasoning is formally determinate there must be a designer in virtue of whom our understanding can have formal determinacy, and this designer is best conceived of as God, who is immaterial. (Here's where Dennett and the ID-designer fall off.) But this is futile. Any formal determinacy would still be *extrinsic* and *derivative* to the understanding. We would be approximating or simulating formal understanding rather than actually possessing it. Nothing that we actually do could be formally determinate *in se*. Any formal determinacy would be superadded and sustained wholly from without, an extrinsic and derived feature of our abilities and activity. Our mathematical abilities and activity, for example, would be mathematically determinate in virtue of something other than the very

abilities and activity in question, say an intended interpretation or model—even if divine—imposed from without and *in principle* interchangeable for an incompossible one (say, if God wills or had willed otherwise). Our practices would be in themselves wholly compatible with incompossible pure functions or formal rules. Furthermore, how can we access God's model for interpreting our "formal thinking"? How can we gauge whether one has acted in accord with a formal rule or has violated it if the formal rules to be followed are determined by the intentions of God? Who has (infallible?) access to the intentions of God? I am not saying that the ultimate explanation of our capacity for formal understanding may (or may not) lie in some divine cause or intervention. All I'm saying is that if it does, it cannot be by extrinsic determination and denomination. Understanding is, so to speak, truly ours!

4. RETREAT FROM PEOPLE

Ross anticipates something analogous to Dennett's objection. Here is his formulation of what I will call the Quine-Dennett (Q-D, for short)[234] objection:

> Humans do not really add, either; we just simulate addition. Pure addition is just as much an idealization for us as is $E = mc^2$. We can *define* pure functions but cannot *perform* them.
>
> (Ross, 2008, p. 121)

> We do not really add, either; we just simulate addition. Pure addition is just as much an idealization as $E = mc^2$. Of course, we can define such pure functions but cannot realize them; that is just a case of the many functions we can define which cannot be computed by any finite automaton, or any other computer either. In a word, the fact that there is no pure addition and no pure conjunction or *modus ponens* is no odder than the fact that there are no perfect triangles.
>
> (Ross, 1992, p. 145)

If formal understanding is (the sole consequence of the actualization of) a purely physical capacity, then human beings cannot have capacities and abilities, nor can they engage in activities, that are formally determinate. For nothing material can on account of its materiality be determinate in pure form. The Q-D objector concedes this last point, but just goes on happily to apply it to human

---

[234] Perhaps this objection should be called instead the Churchland-Dennett objection, since the Churchland's position notoriously implies that there cannot be such a thing as mental properties (e.g. conscious experience, intentionality, etc.) as common sense understands them (though they might have some utility in our everyday life) but only brain processes. Indeed, realism about the psychological is untenable since the psychological is but a useful illusion or what have you. To put it differently, all there *really* and *objectively* is to human psychology is neurobiological processes and mechanisms. Hence the name eliminative materialism. "Eliminative materialism is the thesis that our common-sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience. Our mutual understanding and even our introspection may then be reconstituted within the conceptual frame-work of completed neuroscience, a theory we may expect to be more powerful by far than the common-sense psychology it displaces, and more substantially integrated within physical science generally" (Churchland, 1981, p. 67, cf. 1993, 2007, 2013). Given KRA, if Churchland is correct, there cannot be such a thing a genuine, real, and objective formal understanding. But Quine developed the indeterminacy of translation and championed epistemology naturalized, and since we've already discussed the conceptual interconnections between Quine's indeterminacy considerations and Kripkestein's rule-following skeptical paradox, and Dennett himself names Quine as amongst those philosophers in agreement with him on original or non-derived intentionality, it seems appropriate to formulate the objection as one that could be taken up on his behalf by some like-minded philosopher.

beings. On this view, then, we have only derived formality. If anything, we could only simulate or approximate what is called 'following logical and mathematical rules', 'carrying out formal operations', 'grasping and defining pure functions', etc.[235] But there are several problems with the Q-D objection, some of which prove lethal.

4.1 THE ARGUMENT FROM CERTAINTY

Saying we only simulate pure functions implies that we do not, and cannot, *really* carry out and understand pure functions. To understand a pure function is, in part, to acquire the ability to do certain things such as e.g. to explain it (say, by demonstration) and carry it out. But if we do not *really* understand pure functions, we cannot *really* grasp, define, or explain any, and if we only simulate their application or computation then that means we do not *really* apply or compute them. We would not able to do anything that is formally determinate. We could only construct concatenations of linguistic symbols which we, unfortunately, could not even come to *see* as a formal definition—i.e. an act which could only be described as a simulation of a formal definition (whatever that means). For the ability to perform a pure function springs forth from the same capacities that make it possible for us to acquire the ability to define it[236], and one cannot define a pure function if one cannot determine what it is to carry out that function.[237] We can only *simulate* understanding, explaining,

---

[235] "[…] to deny that our judgments are of definite logical forms and pure functions conflicts with our own certainty and with what we tell our logic, mathematics, and linguistics students about validity, proof, and formal syntax, and leaves us unable to explain what we do when we do mathematics, logic, or any other formal thinking." (Ross, 1992, p. 137) "…to deny that our judgements are of definite logical forms—namely, pure functions—conflicts with our accounts of validity and of logical transformations and leaves us unable to explain what we do when we do mathematics, logic, or any other formal thinking or even simple conjunction." (*ibid.*)

[236] Obviously, one can learn how to add without having the slightest clue as to how to define it. Again, there are different degrees and kinds of understanding. But then again, there is no such thing as one single way of defining a pure function like addition. For definitions in mathematics are, fascinatingly enough, possible across different mathematical fields. Thus, it is possible to define addition algebraically, geometrically, topologically, and set-theoretically, for example. Now, most of us learn to define functions like addition by means of ostensive definitions and exemplifications of what it is to add. The point is that one can know what addition is and know what it is to add, without knowing how to define addition in a non-demonstrative or non-exemplificative way. But one cannot be said to be able to define addition (not writing something down that just happens to be a definition of addition, but actually defining it) without being able to add—in the sense of not knowing what it is to add.

[237] Only something capable of understanding a pure function could acquire the ability to define a pure function. The same goes for grasping, explaining, performing, realizing, computing a pure function. Now, by 'defining' I do not mean providing a definition in the manner in which someone inside Searle's Chinese room could be said to provide definitions. Suppose the person inside the Chinese room is asked to define 'differential equation' (perhaps we should call it the 'Formal Room' instead), and an appropriate response is outputted from the room. It remains that the person inside the room (assuming she's total ignorant as to what she is being asked to define) would not have defined anything, even if she has provided a definition mechanically, that is, even if what she constructed merely out of the room's program rules can be called a definition of 'differential equation' (e.g. "any equation which contains one or more terms involving derivatives"). For she has no understanding as to what she's being asked to define. In following the program rules or instructions the person is not *defining* anything, but mechanically constructing something that can be called a definition of 'differential equation'. But this requires no understanding of what a differential equation is. The same goes if she is asked to solve a particular differential equation. She does not have the ability to do so, whether in Chinese or any other language. Yet, being able to follow the rules whose application generate a formulation or definition of 'differential equation' requires that she understand those very rules (that is, that she understand what it is to apply or follow those rules). Otherwise she cannot apply them—except by force or by having some other external agent move her hands in such a way that she writes down the definition of a differential equation. But the person inside the 'Formal Room' does understand and therefore apply the formal, mechanical rules even if she cannot understand or know what the result of their application output is/means. As we've said, there are different levels, degrees, and kinds of formal understanding. Our argument applies just as much to the person inside the 'Formal Rule' as it does to the person who understands, in

defining, and performing pure functions. But that we in fact can and actually do understand and carry out pure functions like addition "is more certain that any argument to the contrary," (Ross, 2008, p. 121) and anything to the contrary is almost as unbelievable as the suggestion that we aren't *really* reading these words, but only "simulatively" so. "It denies outright that we do the very things we were sure all along that we do. It's like saying we can't read these very words." (Ross, 2008, p. 121) It's practically saying we cannot understand mathematics, logic, and formal linguistics, nor any of their applications. Furthermore, no one could be sure to understand philosophical reflections such as this one, not even the Q-D objection, since no one could understand such a thing—including Quine and Dennett. "[…] to deny that our judgments are of definite logical forms and pure functions conflicts with our own certainty and with what we tell our logic, mathematics, and linguistics students about validity, proof, and formal syntax..." (Ross, 1992, p. 137)

4.2 THE ARGUMENT FROM EXPLANATORY SUCCESS

"[…] to deny that our judgments are of definite logical forms and pure functions … leaves us unable to explain what we do when we do mathematics, logic, or any other formal thinking." (Ross, 1992, p. 137) Moreover, it leaves us unable to explain not only what we do when we do theoretical (pure mathematics, logic, formal linguistics, philosophy of mathematics, etc.) and practical (computer science, engineering, physics, and so on) sciences but also the respective successes that we've had in such disciplines. It also threatens to undermine the rational reliability of such disciplines. "Yet our doing these things [e.g. thinking in formally determinate ways] is essential to *the reliability* of our reasoning" (Ross, 1992, p. 146; my italics), and (I would add) to the reliability of such sciences and disciplines. It is hard to see how theoretical and applied sciences, which are human practices that depend on the understanding of pure formal structures, would be so much as possible if we couldn't understand pure formal structures.

4.3 THE ARGUMENT FROM COMMONSENSE

Finally, the Dennett objection implies something that is extremely implausible (and self-undermining, as we will later show): namely that we do not and cannot ever *really* reason logically and mathematically. All we do is simulate logical and mathematical reasoning. But then none of our arguments can ever be really valid (or invalid), including Dennett's arguments in favor of indeterminacy. "The final and greatest cost of insisting that our judgements are not more determinate as to pure functions than physical processes can be, is that we can do nothing logical at all, and no pure mathematics either. Now, who believes that?" (Ross, 1992, p. 146)

These 'soft' (but by no means trivial and harmless) considerations threaten to do away with the rational reliability, acceptability, and justification of the enterprise at the service of which such objections claim to be: science. There would no such thing as scientific knowledge, since it would not be possible to understand anything scientific insofar as scientific knowledge depends on formal understanding. There would be no such thing as mathematical and logical knowledge, since no one could *really* understand logic and mathematics. Therefore, there can be no such thing as a *real* practice of mathematics and logic, because no human being can *really do* mathematics and logic. This is expensive. Consequently, no human practice can consist in the *real* comprehension, apprehension, and execution of mathematical and logical functions. The conglomeration of formal practices that

---

addition, what a differential equation is. KRA shows that such kinds or levels of formal understanding cannot be wholly a matter of matter.

fall under the heading of mathematics and logic, and those that are foundationally constituted by mathematical and logical understanding, would be no more than simulations (of what?) or approximations: nothing but systematic and organized simulacra.[238] Naturally, we are left without an explanation of the successes of mathematics and logic, and of their extensions (e.g. formal linguistics and theoretical computer science) and applications (e.g. physics, chemistry, and engineering). I gather some philosophers would be more than willing to bite the bullet and do away with such disciplines and our ability to do them. But one cannot coherently and intelligibly bite the bullet and do away with such abilities and activities.

4.4 THE ARGUMENT FROM NON-SIMULATION

A *simulation* is consequent upon and unintelligible without what it is a simulation of (that is, the non-simulated case), and any simulation of the realization of a pure function will be consequent upon and unintelligible without the non-simulated realization of a pure function. A simulation is what it is only in virtue of that which it simulates.[239] But since, according to the objection, we can have no understanding of that which the simulation is supposed to simulate (e.g., recreate, imitate, approximate, or what have you, a formal operation) we can have no understanding of what it is to simulate such a thing, and thus no intelligible notion of the simulation of a pure function. In other words, if the Q-D objection were correct, we couldn't even tell what our intellectual activity is a simulation *of*, let alone that it is a simulation. To say we only simulate pure functions or formal operations but never really compute them or carry "them" out, without being able to understand what is being simulated and what it is to genuinely compute "them" or carry 'them' out, involves a conceptual mistake.

Perhaps the Q-D objector will respond by simply conceding that we cannot even understand what it is to simulate a pure function, since such understanding is parasitic on understanding pure functions and on what it is to do them—something which, according to the Q-D objector, we cannot have. We can understand what simulating other things are (presumably, only non-mathematically based simulations like real life earthquake simulacra)[240] and on the basis of that go on

---

[238] There would be no such thing as languages since no one could grasp and follow grammatical rules if grammatical rules are, even if partly, formal rules or if grammatical rules make possible the generation of formal rules and formal languages. But to do away with language? We will deal briefly with linguistic understand and Chomsky's approach to language in Chapter 6. Here we will simply confine ourselves to logic and mathematics, i.e. to the capacity and abilities for logical and mathematical activity, and show why the Dennett objection cannot be coherently held with respect to these.

[239] I am talking about the simulation of human *activities*, i.e. of the exercise of intellectual abilities. 'Simulation' is a polymorphous concept (e.g. consider 'simulation' as it applies to an Earth quake simulacrum at a school and 'simulation' as it applied to computer simulations of the trajectory of an airplane).

[240] The Q-D objection makes it impossible for anyone to understand mathematized simulations like computerized weather simulations, flight simulations, market crash simulations, chess-player simulators, etc. Indeed, the objection seems to make such a thing as "mathematized simulation" impossible not only as something *understood* but also something *made* or *produced* since computational simulations are the product of practical formal understanding. An explanation is in order, however, about the difference between the concrete implementation of computations and modelling computationally physical phenomena if both are simulations. "Questions on the nature of concrete computation should not be confused with questions about computational modeling. The dynamical evolution of many physical systems may be described by computational models. Computational models describe the dynamics of a system that are written into, and run by, a computer. The behavior of rocks — as well as rivers, ecosystems, and planetary systems, among many others — may well be modeled computationally. From this, it doesn't follow that the modeled systems are computing devices — that they themselves perform computations. Prima facie, only relatively few and quite special systems compute. Explaining what makes them special — or explaining away our feeling that they are special — is the job of an account of concrete computation." (Piccinini, 2017)

to say, by analogy, that whatever simulating a pure function is, that *is* what we do—even if we cannot say, specify, and understand what *that* is. Similarly, one can go on to say that whatever realizing a pure function is, that is what we do *not* do. But this will not work, partly because of the reasons to be given in (5) and partly because of the following. If one cannot understand a pure function, one cannot understand what it is to realize or perform a pure function, and therefore one cannot understand what it is *not* to realize or *fail* to carry out a pure function. If one cannot know what it is to add, one cannot know what it is *not* to add—in the sense of what it is to be unable, and what it is to fail, to follow the rule of addition or follow a rule other than addition. There cannot be any basis for saying "whatever such-and-such formal operation is, we do not do it" if we cannot understand and therefore come to know what it would be to do such-and-such a formal operation, and thus what it would *not* be to do such-and-such formal operation. For, one can only intelligibly say that one cannot do some formal operation *x* if one knows what it is to carry out such a formal operation *x*, and consequently one cannot say intelligibly that one cannot carry out some formal operation *x* unless one knows what it is to carry out (and what it is to not carry out) formal operation *x*. In short: if one cannot understand a pure formal operation *x*, then one cannot understand what it is to carry out (correctly or incorrectly) *x*. The same goes for simulating *x*: if there is no intelligibility to the notion of the simulation of a pure function, there cannot be any intelligibility to the claim that we simulate a purely formal operation. To say, "whatever simulating *x* is, that is what we do" will not work because the objection grants that we cannot understand what "that" is, and therefore we have no criteria for determining what it is and what it is not to do "that" (that is, what it is to simulate and what it is not to simulate *x*) and thus no criteria that determines whether we do it or not. Likewise, the objector's claim "whatever *x* is, we do not do it" makes sense only if it can be known or understood what it is to do, and *not* to do, *x*. But did is precisely what the objection denies. The objection is sheer nonsense (see 4.5 and 4.6 below).[241]

In addition, the criteria for what it is *not* to follow a certain rule (say, to produce 5 for any pair of numbers greater than 57 is *not* to add those numbers), i.e. what it is to fail at it, to do something other than it or to carry it out incorrectly, depend upon the criteria for what it is to follow that rule, i.e. what it is to accord with it or to carry it out correctly. If we cannot come to grasp the latter, we cannot come to grasp the former. The exact same considerations apply to simulations, for there are successful and unsuccessful simulations. The criteria for what it is *not* to simulate something successfully is dependent, in part, upon what it is to simulate it successfully (it is also dependent on what is to be simulated, otherwise there is no telling when the simulation succeeds). But if there are no criteria for simulating successfully and adequately, there are no criteria for simulating unsuccessfully and inadequately, and there would be no intelligible basis for stating of anything that it simulates something successfully, or that it does not. So the objection collapses.

Ross tells us, if all we did were simulations, we "couldn't really do logic or mathematics or any other formal thinking; we couldn't even castle in chess, but could only simulate it, without there being any explanation for what "it" is, *since thinking so as to define it would be to perform a pure function. The relation of simulation to a mathematical or logical operation will, itself, not be definable without the prior notion of pure functions*." (Ross, 2008, p. 122; italics are mine) The concept of simulation is a *consequent* or

[241] The objector can still claim that we just can't know whether we think in ways determinate in form *and* whether we simulate thinking in ways determinate in form. In the case of pure functions, we are unable to distinguish realization from non-realization and simulation from non-simulation. Consequently, we are unable to differentiate a realization from a simulation since we cannot acquire the understanding that would enable us to determine what counts as a realization (and what not) and what counts as a simulation (and what not). But then there is no intelligible basis to the Q-D objection, which hinges on there being a difference between realization and simulation. There's also the problem of explaining what it is to *know* that it is not possible to differentiate the realization and simulation of pure formal structures if there's no such thing as understanding what pure formal structures are.

dependent concept, and when used in relation to logical and mathematical operations, it will not be definable nor intelligible without the prior notion of the logical and mathematical operations themselves. One cannot know what a simulation is, or whether something is a simulation, if one does not (or cannot) know what it is a simulation *of*, i.e. what it is simulating, or what it is to be a real, non-simulated instance of what is being simulated. Otherwise, there would be no distinguishing a simulation from a non-simulation, and no basis for saying that something *is* a simulation. To be able to know whether something simulates, say, the act of adding one needs to know what it is to add (what *adding* is). Similarly with "approximation".[242] One cannot know whether something physically approximates a Euclidean right triangle if one does not (or cannot) know what a Euclidean right triangle is, and there cannot be a simulation of a Euclidean triangle if there is no such thing as a Euclidean triangle (whatever its ontological status).[243] Simulating the act of adding or of conjoining already presupposes that there is such a thing as adding or conjoining, and our assessment of something as a simulation of adding and conjoining presupposes that we know and understand what it is to add and conjoin. The Q-D objection is therefore a *petitio*. Simulation is consequent upon, and presupposes, non-simulation. The simulated case presupposes, and is intelligible without, the non-simulated case. And with respect to pure functions, the notion of the simulation of a pure function will be neither definable nor intelligible without the prior notion of pure functions.

Finally, it is still the case that "a simulation, as far as it goes, can always be a case of something else; so if that were all we could do, it would never be determinate what we are doing at all." (Ross, 2008, p. 122) Any case that counts as a simulation of the realization of ψ, where ψ is a pure function, will just as well count as a simulation of the realization of ψ^*—a pure function incompossible with ψ. And as we've shown already (chapters 2 and 3), fixing which pure function a physical process is a simulation *of* is a matter of an interpretation,[244] the correctness of which is in turn fixed and

---

[242] I am talking about approximation as it applies to the physical in relation to the mathematical and logical, not about approximation as it is predicated of one mathematical structure in relation to another mathematical structure. Take, for instance, the decimal representation of √2, which is '1.41421356237309504880168872420 97…' This an instance of an *inexact* decimal representation. All decimal representation of irrational numbers like since √2, will be inexact and without any discernable or predictable pattern of digits. Following an algorithm for calculating each next digit, we can compute more digits and write them down. The more digits we compute and write down the "closer" we get to √2 without ever really getting to √2. (The notion of approximation is represented symbolically as "≈"—which means "is approximately". We can then go on to express symbolically the same proposition: '√2 ≈ 1.41421356237309504880168872420 97…') So we say that the relevant decimal "approximates" or "is an approximation of" √2 (we do *not* call this a simulation, though). Alternatively, we say that "√2" is approximately that decimal. But this is approximation between formal objects and it is something we decide to call as such. The physical approximation of a formal object or structure is categorially different because it is a "relation" between structures of different ontological orders, each of which is determinate and determinable in ways the other cannot be, the former being embedded in conceptual dimensions that cannot be intelligibly attributed to the latter, and conversely.

[243] Can one simulate fictions or fictitious entities? I guess so. Though I'm not sure there is such a thing as simulating a "unicorn", though things might be different with fictional scenes and events, but there is such a thing as representing them in some way or other (say, by constructing a carrusel made of unicorns rather than horses). So, one can ask, what about amusement parks? Don't they simulate fictional events and stories? Think of the old famous attractions of Universal Studios (e.g., Back to the Future, Jaws, King Kong, Terminator, etc.) What about fictional characters? If I put on a Micky Mouse custom, and try to talk and behave like "him", I'm surely imitating or simulating him, don't I? That the simulated is dependent on the non-simulated for its being as a simulation does not mean that what is simulated is *real*. Amusement parks simulate movies or fantastic worlds all the time. Ontologically, conceptually and epistemically, a 'simulation of *x*' is dependent and consequent on '*x*'. The interesting point emerges when we come to act and activities, and in particular, formal acts and activity which seem anything but fictional (like our arguing for KRA thus far).

[244] It is important to clarify that I am concerned with those physical mechanisms that are generally taken as capable of engaging in the sort of activities that human beings engage in (say, following logical and mathematical rules and carrying

determined ultimately by the choice and intentions of the designer or observers.[245] Simulation, like approximation, of the mathematical and logical by the physical is an observer- or mind-dependent feature of physical things. "A machine simulates adding, calculating, recalling, and so on. What it does gets the name of what we do because it reliably, even more reliably, gets the outputs we do when we add and can get those outcomes for cases we cannot efficiently process. The machine adds the way robots walk and use tools. The names are analogous, but not merely equivocal or metaphorical. The machine attains enough reliability, stability, and economy to achieve realism without *that* reality. A flight simulator has enough realism for flight training; you are really trained, but you are not really flying. The same holds for Big Blue and chess." (Ross, 2008, p. 120)[246]

4.5 THE ARGUMENT FROM PRAXIS: PRAGMATIC CONTRADICTION

For one to go on to assert that one does not think, judge, or reason in any formally determinate way one has to think and judge with all the formal determinacy that is being denied is possible. Thus, to deny that we really add, conjoin and carry out recursive operations, for example, requires

(a) that one judge in a way that is determinate amongst incompossible pure forms, otherwise it would be indeterminate what is being done (e.g. negation, "$\neg\psi$", as opposed to some other incompossible function, "qu-negation"),[247] and
(b) that one understand pure functions (i.e. have formal understanding), otherwise it would be indeterminate what is being denied.

Indeed, if no one can understand pure formal structures, no one can understand Kripke's argument (not even Kripke himself) and so no one can accept or reject it, except ignorantly. So in order to even be able to study, grasp, explain, and assess Kripke's argument one must be able to understand pure forms like addition and quaddition and think in formally determinate ways. Yet the objection

---

out formal operations) when doing mathematics or reasoning logically. There are, of course, computational simulations of physical processes and mechanisms, and a computational simulation need not be a simulation of something computational (e.g., weather simulations). So there are different species of simulations. The question I am addressing is whether physical mechanisms, and their processes, can be capable of formal understanding and thus of possessing the set of abilities whose exercise constitutes real formal operations. So I am concerned with the claim that physical mechanisms qua physical can do *the very same* thing that human beings do, i.e. that "computation" in humans is univocal with "computation" in physical computing mechanisms. I say no. The former can only count as a simulating, or approximating, the latter. However, even in the case where it is the physical mechanism or process that is being simulated by a mathematical object, rather than the other way around, formal indeterminacy will abound. The same physical mechanism or process cannot exclude quus-like mathematical simulations.

[245] Again, this is not a triviality argument. The point is not that the physical object satisfies any and every computational interpretation; rather, the point is that even if the system does not satisfy any and every interpretation, and in fact excludes an indefinite range of interpretations such that we can reliably say that it does not compute any and every function, it is still the case that the physical object cannot exclude incompossible functions from being realized and implemented at the same time and in the same respect. There will always be at least one other function not compossible with the intended function which the system can be said to implement or realize at the same time and in the same respect.

[246] Auto-pilot is an interesting case but, again, fundamentally no different. The computerized model achieves enough reliability to achieve the output-reality without the relevant conceptual, perceptual, and proprioceptive skills that would make that an otherwise human activity. The pilot and the auto-pilot do not *do* the same thing, but they can both achieve more or less the same outcome in terms of the plane's trajectory and such. But the auto-pilot mechanism can be made to conform to incompossible formal descriptions.

[247] This true of the reasoning that leads to the Q-D rejoinder: If the physical is formally indeterminate and human understanding is purely physical, then human understanding is formally indeterminate. The physical is formally indeterminate and human understanding is purely physical; therefore, human understanding is formally indeterminate.

denies that we can ever do this. This is a pragmatic or performative contradiction, for the very act of formulating the objection consists in doing the very thing the objection theoretically says cannot be done.[248]

As Ross points out, the objector can always respond by rejecting (b), arguing instead that "one can indeterminately deny yet include what is determinate, as when one might say, 'no one can really add' without being specific as to what is denied, in effect meaning "do anything *you* would mean by 'add,' " without supposing there is such a thing." (Ross, 2008, p. 199, n. 23) This is a strange objection, and Ross does not attempt a refutation of it. But it does not work. Firstly, the objector is still negating a proposition—which makes her judgement *formally* determinate, even if *semantically* indeterminate. Indeed, her judgment has to be determinate in form, otherwise there would be no fact of the matter as to what she is doing and the whole objection would collapse, since no such thing as a negation could take place. If (b) does not undermine the objection, (a) will. But secondly, if by "no one can really add" the objector means to deny whatever someone else means by "to add" without herself understanding what "to add" is, then either *that* someone else understands what is "to add", in which case someone *does* understand a pure function and there is such a thing as formal understanding (viz., the person who denies would be relying on someone else's formal understanding), or that person is likewise using it to mean what someone else means by "to add" and so on *ad infinitum*, in which case nothing at all is being denied (or something altogether different is). But to deny nothing, or to deny something altogether different, is not to deny what is in question: understanding addition—or any other pure function for that matter. Hence, the latter horn of the dilemma shows the objection to be a red herring. In addition, it is hard to see how responding that one can deny something without knowing and understanding anything about what is being denied really helps, dialectically, the objector. For it amounts to her conceding that she is denying something she is woefully ignorant about and incapable of understanding. But at any rate, to deny that we do something while also holding that we cannot know what that something is *makes no sense*, since to know whether we do something or not requires that one understand what it is to do that something and one cannot determine whether we do something if one cannot understand what it is to do it.

4.6 THE ARGUMENT FROM INTELLIGIBILITY

This is an extension of the last point. We say that the objection is altogether theoretically incoherent. Let us assume that the Q-D objector is correct: human beings cannot have formal understanding—alternatively, intellectual capacities are not formally determinable; therefore, human intellectual abilities and activity cannot be formally determinate. If the objection were true, it would require that humans could not grasp *it*. We would not understand the objection. To grasp, formulate, espouse, and defend an objection in favor of the formal indeterminacy of human understanding would be impossible. If the objection were true, the Q-D objector would not be able to comprehend, explain, express and defend it. In fact, if the objection were true, the objection could not be intelligibly made. For no one—Quine, Dennett, or anyone else—would have the capacity for coming to understand what 'being formally indeterminate' is! For, again, a grasp of formal (in)determination is parasitic on a grasp of definite formal structures and rules, which is what

[248] This is the exact same reason why the skeptic's argument (Kripke, 1982) *qua* full-blown skepticism about rule-following fails. For in order to be able to make and grasp such an argument one has to do the very thing the argument says one cannot do, think in formally determinate ways—e.g. as when defining quaddition and distinguishing it from addition. One must understand pure functions, e.g. addition and quaddition; otherwise the argument cannot get off the ground. The very argument is an exercise of, and manifests, formal understanding. For a similar point, see (Hoffman, 1985).

the objection denies we can have. But then there cannot be such a thing as *objecting* to the formal determinacy of human understanding or *arguing* for the formal indeterminacy of understanding. No one could do such a thing. Hence, there cannot be such a thing as the Q-D objection. The statement "human understanding is formally indeterminate" if true *should not be intelligible to any human being* since it would not be possible for anyone to understand what formal (in)determination is, and any argument in support of it *could not be made nor followed by any human being* since it would not be possible for anyone to think in the requisite formally determinate ways.[249] The objection entails that there cannot be any intelligibility to it.

Additionally, one can always argue in a Wittgensteinian fashion by insisting that in order to say what counts as formally indeterminate, one has to be able to say what counts as formally determinate. That is to say, to differentiate between simulating the following of a formal rule and really following a formal rule one must be able to grasp genuine cases of (following) the rule, or at least what it would be to really follow it, which can serve as a paradigm upon which the distinction can be intelligibly built (it's like the distinction between what is real and what is fake, e.g. real vs. fake diamonds: the distinction is unintelligible, that is, nonsense, if there are no criteria that establish what it is to be *real*).[250] Otherwise, no one could distinguish one from the other and the whole skeptical argument would collapse. So the objection itself presupposes, or requires, formal understanding.

4.7 LESSONS FROM THE Q-D OBJECTION

[249] This has nothing to do with the philosopher's idea of truths that we cannot *in principle* come to know, and of there being some other beings (say, some hypothetical higher species) that do or could have access to truths about us and about the world that are *in principle* inaccessible to us (see e.g. Nagel, 1986). For my argument is about what *is* "accessible" to us, about what we *can* and *do* understand. The Q-D objection itself is a much a proof as any that formal truths and rules are within our cognitive reach. One cannot understand what formal (in)determinacy is if one cannot have formal understanding. But to understand a formal rule is, in part, to know what it is to follow it, and thus to be able to follow it. It is to have the ability to follow or apply that rule to whatever degree of competency one has (e.g., to understand the addition function is, again, to know what it is to add; to understand conjunction is, in part, to know what it is to conjoin; and so forth). But then one cannot understand anything about formal determinacy or formal indeterminacy unless one can grasp and think in formally determinate ways.

[250] I am talking about criteria for intelligibility, not criteria for empirical truth. Whether a given diamond is a real one or a fake one is a matter of empirical investigation, thus requiring empirical confirmation or disconfirmation. But the reason why the issue can be empirically investigated is because the distinction between fake diamonds and real diamonds *makes sense*, that is, is an intelligible distinction. There *is* such a thing as what it is to be a fake diamond, as opposed to a real diamond. But if there were no such thing, that is, no intelligibility to the distinction, nothing could be empirically confirmed or disconfirmed since the distinction would make no sense and would express no distinction at all. If the distinction between a fake diamond and a real diamond made no sense, there would be no such thing as *what it would be for something to be a fake diamond* and *what it would be for something to be a real diamond*, and thus no amount of empirical investigation could determine whether a given diamond is real or fake since there would be no criteria for being real and for being fake, i.e. no criteria for what it would be to count as fake or as real (unless the community of investigators invent them or agree upon a definite criteria; but that just confirms our point). Intelligibility is prior to, and a necessary condition for the possibility of, empirical investigation (and *any* kind of investigation). Let me also note that this distinction between what is real and what is fake is, of course, a plastic one and will vary depending on which kinds of things are being talked about (consider the stark differences in criteria between 'real vs. fake diamonds', 'real vs. fake intellectuals', 'real vs. fake friends', 'real vs. fake advertisements', 'real vs. fake money', and so on). Yet in all such cases it remains that, in general, criteria for what it would be to be a case of *x* is prior to, and necessary for, being able to 'go out and verify' what or who *actually* satisfies *x* and what counts as 'faking' or as not being a genuine case of *x*. So the general point remains: to *understand* what it is to be 'fake' in a given domain (e.g., money) one must understand what it is to be 'real' in that domain, and that whatever is fake is unintelligible without the conceptually prior notion of what is real.

All of these considerations, which are my own elaborations of Ross's own objections to the Q-D objector, imply that the human intellect, and in particular formal understanding, cannot be (wholly) physical. The viewpoint that human intellectual activity is formally indeterminate, which is what would be the case if the intellect we a purely physical capacity, cannot be consistently and intelligibly held. It leads to absurdity and nonsense. Therefore, intellectual activity (i.e. logical reasoning, assessing the congruence of triangles, grasping and applying syntactic rules, carrying out formal proofs, etc.) is formally determinate and cannot be (wholly) physical.[251]

> If we can agree that either (1) we do have such definite thought processes as I described, cases of conjunction, determinate among all incompatible functions, and that they cannot wholly be physical processes (or functions among physical processes only), or (2) we never perform such processes but at most simulate them, then I am content. For I shall then wait for the counterattack to support (2), the one that explains the status of all those functions I cannot really perform and only think I can define (for to define one is to perform another one), and, in particular, explains the success of mathematics and pure logic, especially natural deduction systems and the proofs of completeness of propositional calculus, and offers a worked-out contrast between adding (which no one, apparently, can do) and simulating adding.
>
> (Ross, 1992, p. 148)

Before considering other objections that can be raised against KRA we will first address some corollaries to Ross's argument for the formal determinacy of human understanding. The following related but distinct considerations must be taken as *addenda*. Our main argument, KRA, stands even if the following fall.

5. THE PARALLEL ARGUMENT AS TO CONTENT

Recall that at the beginning of this work (chapter 1) we made a distinction between the content of understanding and the forms of understanding. I called the former, perhaps misleadingly, the *semantic* component of understanding, and I called the latter the *formal* component of understanding. The semantic component is what our thinking, reasoning, judging, inferring, calculating, or what have you, is *about*. The two, though conceptually separable, and are not so in reality.[252] I have already said that I'm not of the opinion that one can have one without the other, that is to say that the *capacity* for understanding formal rules is not the very same capacity for following or applying formal rules, and vice versa.[253] For, as we've already said, to understand a

---

[251] As I've said before, if an activity *x* cannot be wholly physical then neither can the ability to engage in *x* be wholly physical, since the activity is the ability exercised. Likewise, the capacity for acquiring the relevant ability cannot be wholly physical either. I call the innate capacity human beings have for formal understanding (and understanding in general) "the intellect".

[252] Even if, *à la* Chomsky, they are in reality separable (and, furthermore, innately separated), it does not affect our point. Chomsky believes that some formal structures (namely, those of universal grammar) are applied and followed but are not understood (and perhaps need not or cannot be understood). They are 'tacit', which is another way for unknown. So the rules of universal grammar are tacit and tacitly applied, and therefore unknowingly applied. Though I take Chomsky's position on rule-following and rule-understanding to be wrong-headed, it is irrelevant to the point at hand. If we concede for the sake of argument that Chomsky is right, we can just rehash the argument with respect to the capacity and ability to follow formal rules and ignore whether knowledge and understanding are necessary. Our argument would still make the naturalization of that capacity and ability impossible.

[253] Note that I am talking about the *capacity*, not the ability (though the same can be said with respect to abilities). I mean that to be the sort of being that can come to acquire understanding of pure formal rules is to be the sort of being that

formal rule is in part to come to know what it is to follow it and to have the ability to follow it.[254] Be that as it may, the interesting feature is that there is a parallel argument as to the semantic or

---

can come to acquire the ability to follow pure formal rules, and vice versa. The capacity for formal understanding is at least in part the capacity for following formal rules or carrying out formal operations. Intellectual capacities are formally determinable, but not formally determinate (*pace* Chomsky, for one cannot follow rules one cannot or does not understand). As to the ability, that is, the acquired skill to carry out a definite formal operation, we say that it is that which is formally determinate 'in potency' when not exercised but formally determinate 'in actuality' when exercised, much like my ability to write this thesis in Spanish is determinate but in potency as I have not actually written it in Spanish. At any rate, all of this is not necessary for KRA to work. If one does not believe that formal understanding consists, at least in part, in the possession of certain abilities (say, like having the ability to add negative integers, to apply the rule of distribution in propositional logic, to use natural logarithms, etc.) relative to that which has been understood, you certainly cannot intelligibly deny that we do, and can, have such abilities and actually do, and can, exercise them (ignoring accidental impediments and factors). We do prove theorems and meta-truths about formal systems, we do solve complex equations, we do infer deductively in accord with determinate valid forms of inference, we do define vectors, and so on. The argument could've well been dubbed "the argument against naturalizing formal activities (and, therefore, the ability or set of abilities for carrying out such activities)". Thus, even if one denies that coming to understand a formal rule or pure function implies (or consists in) the onset of a set of abilities, in particular the ability to apply and follow the rule, KRA remains unscathed. For, all that KRA requires is that we be the sort of being capable of acquiring the ability to follow and apply pure formal rules. And since to actually follow and apply pure formal rules requires features and abilities no material thing and process can have (formal determinacy, as our Kripkean argument has shown) and given that we actually *do* (and can) follow and apply pure formal rules in a formally determinate way, it follows that the activity that ensues upon the exercise of that ability (i.e. the ability to carry out formal activities and practices) cannot be wholly physical, and neither can the ability, and innate capacity, for such an activity be wholly physical. Likewise, if one denies that knowing what it is to follow a formal rule or compute a pure function does not imply that one is able (that is, has the ability) to add, it does not follow that KRA is unsound or misconceived or what have you. Though this is a bizarre position, all that suffices for our argument is that we *can* and *actually do* acquire such an ability (say, the ability to add) whose exercise constitutes a determinate kind of activity (namely, adding). Since adding is a formally determinate activity, the ability to add is an ability to carry out something no material thing can possibly carry out. Hence, formal activity and the underlying abilities are not (wholly) physical. The implication will be the same regardless. It is, however, my opinion that to suggest that one can have the ability to follow a formal rule without understanding the given formal rule is sheer nonsense and involves a conceptual mistake, just as much as the denial that understanding a formal rule is, in part, the ability to do certain things with respect to the rule. Otherwise, someone's following a rule will be much like a parrot's squawking an English sentence: the parrot may squawk but it does *not speak*, and is not *speaking*, English. (A parrot cannot engage in *speech*, and thus cannot really *say* anything; it can only squawk sounds that mimic speech, and it doesn't speak English because it neither means nor understands what it squawks. It cannot follow any linguistic rules. See e.g. Hacker, 2005, p. 360.) Without understanding there is no differentiating mere conformity with (i.e. behavior that merely conforms with, mimics, or is *as if* it follows) the rule from *really* following the rule. But then there is no intelligibility to either mere conformity with (e.g. simulation, mimicking, mere coincidence, "as if") and real application of (e.g. actually following) the rule. For the former is consequent upon and undefinable without the latter, but if the latter cannot be differentiated from the former the distinction becomes unintelligible and there would be no such a thing as merely conforming to, *as opposed* to really applying or following, a rule. See my remarks about formal understanding in Chapter 1.

[254] Of course, you can think about, say, differential equations without knowing what it is to solve differential equations nor what a "differential equation" is. Your thoughts about differential equations might have no more content than "the sort of equations physicists work with", or something of the sort. In other words, one's thoughts can be about things one does not comprehend. That much is true. It is a feature of intentionality that one can think about something that one doesn't grasp or comprehend. But then we say that you do not understand, or that you lack understanding of, whatever it is you are thinking about. It is understanding that is of interest to us here: the understanding of, and *ability* thereof to explain, define, solve, teach, etc., such things as differential equations. Of course, one does not need to be able to do all these specific things in order to understand what a differential equation is, though being able to do *some* is necessary. To be able to do *all*, and actually doing them, on the other hand manifests the possession of a deeper sort of understanding with respect to such equations (the understanding of an expert, for example). But if you *do* understand what a differential equation is or have *some* understanding of what such an equation is, then you must at least be in a position (be able) to do something with respect to what is understood, such as being able to solve a differential equation, to see mistakes either upon review or upon having someone else point them out and (depending on the complexity of the problem) correct them, discriminate differential from non-differential equations, and so forth.

intentional component of understanding which leads to the very same conclusion but in relation to conceptual content or linguistic meanings. The conclusion is that no physical thing can be on account of its physicality alone *semantically* determinate. Ross himself calls this argument "the parallel reasoning as to content" (Ross, 2008, p. 123; 1992, p. 150, n. 17), and asks "Can the reasoning above [about (in)determinacy] be applied to content as well as to form?". "The present cases concern the definiteness of the *form* of the thinking. A third, parallel argument can be constructed from the definiteness of the *content* of thought, that thought is definite among incompossible contents in a way no physical process can ever be. Similar underdetermination arguments apply." (Ross, 1992, p. 150 n. 17) Unfortunately, this distinction is obfuscated and ignored by such commentators as Edward Feser (Feser, 2013, 2016) and Dillard (Dillard, 2014, 2011) who treat Ross's argument as if it was solely or mainly about *semantic* determinateness (in particular, the content of formal thinking) rather than *formal* determinateness. [255] We will come back to this later. Here we will look at the argument

[255] Whereas Feser (Feser, 2013) goes back and forth confusedly between formal determination and semantic determination, in later treatments (e.g. blog posts) he makes it clear that he takes the notions of 'determinacy' and 'indeterminacy' to be semantic rather than formal. Formal determinacy is nowhere to be found in Feser's later treatments of Ross's argument. The problem seems to be that Feser conflates the form as the form of some thinking (i.e. conjoining) with the form as the content of some thinking (i.e. thinking about conjunction), and then is misled to think that, like most anti-naturalist arguments in favor of the immateriality of the human intellect, Ross's argument has to do with content or meanings. But this does grave injustice to Ross and turns some of his insights into sheer silliness. For instance, when Ross says, "our thinking, in a single case, can be of a definite abstract form (e.g., $n \times n = n^2$), and not indeterminate among incompossible equally most particular forms…" he does not, of course, mean that whenever we square a natural number the content of our thought is "$n \times n = n^2$". That might be the case in certain occasions as we shift our attention from that which we are trying to square to the square function and formulae, but not always and definitely not necessarily. One does not square the square function; that makes no sense. The content is whatever is being squared, say, a prime number. Ross's claim is, rather, that we act in terms of the pure form "$n \times n = n^2$" whenever we square a natural number. Do not confuse "$n \times n = n^2$" as a form of representation (a particular syntactic-symbolic notation which expresses a rule or function) with the (pure, abstract) form of calculation that "$n \times n = n^2$" represents. The pure form can be represented variously, with different notations, and thus cannot be identified, let alone exhausted, by a particular symbolic notation. Having said that, we need a means of representation. Humans cannot think or understand a pure form without a suitable means of representation, for our cognition is embodied cognition and thus must always be mediated by what the Aristotelians called phantasmata and the incredibly complex sensory and psychological mechanisms that underlie human cognition generally. But not all means of representation are created equal. Some means of representations (e.g., roman numerals) may constitute an obstacle to formal understanding, to insight, and to further developments in formal knowledge. So which means of representation we use matters (see e.g. Lonergan, 1992 on the "image", his preferred label for what we are calling "representations"). We follow the abstract rule that can be expressed with that notation and in doing so our activity is structured in such a way that it can be said to be a case of "squaring a number". But squaring a number and thinking about the square function are not the same thing, for I can think about the square function without squaring any number. There is a basic difference between doing an operation and merely thinking about the operation (a difference but not a strict difference; for, as I've already said, the ability to carry out a given formal operation implies that one can think about that formal operation and vice versa so long as "thinking about" is with a grasp as to what it being thought of). Another example of the confusion to which conflating form and content leads can be found in the following passage by Feser: "Adding, squaring, inferring via modus ponens, syllogistic reasoning, and the like are some of the examples of formal thinking Ross appeals to. Anyone who agrees that material processes are indeterminate in the way Kripke's and Quine's arguments imply but who wants to avoid the conclusion that thought is immaterial will have to deny that any of our thoughts is ever determinate in its content; and writers like Bernard Williams and Daniel Dennett essentially do deny this. But then they will also have to deny that our thoughts are ever really determinately of any of the forms just cited. They will have to maintain that we only ever approximate adding, squaring, inferring via modus ponens, etc." (Feser, 2013, p. 17) Here again adding, squaring, inferring via modus ponens, and syllogistic reasoning, as forms of an activity (an ability exercised) are treated indiscriminately as contents of an act or activity. It is a staple of Feser's appropriation of Ross's argument that it conflates the form of the intellectual act with the content of the intellectual act. We will see this more clearly in §6. It is not, however, entirely Feser's fault since Ross himself talks about "form" and about "formal content", and is quite explicit about his reliance on Quine, Goodman, and Kripke, all of whom were variously concerned with the issue of 'meaning'. So, though the semantic version of the argument runs parallel to the formal one and is indeed a corollary of it,

from semantic determinateness developed by Feser in accord with his reading of Ross which we will subsequently contrast with the parallel argument from content developed by Ross himself.

Consider the *meaning* of words or the *content* of concepts.[256] Feser identifies two features of concepts (or, rather, of the content of concepts?) that distinguishes them categorially from their forms of representations (say, images). Following the Aristotelian tradition, he calls these *phantasms* (Feser 2013, pp. 7-8):

(1) "the *abstract and universal* nature of concepts as opposed to the *concrete and particular* character of images or phantasms", and
(2) "the *clarity and distinctness* of many concepts [as opposed to] the *vagueness and indistinctness* of their corresponding images or phantasms."

Consider our concept of 5,000,001 soldiers, as opposed to 4,999,999 soldiers. Now consider a mental image of the former and a mental image of the latter. They are indistinguishable. The very same image can be said to represent either of those concepts (and infinitely many more, say, the idea of 5,000,003 soldiers). This is a wholly generalizable phenomenon. Hence, a mental image satisfies incompossible concepts.

> We can readily form a phantasm of five trees in a row. But to imagine fifty (not forty-nine or fifty-one) trees in a row will be for most people an impossible task. To imagine five thousand (not more or less) trees in a row is an utter impossibility. But my idea of five thousand or five million trees is just as clear to my intellect as five or ten; I have no more difficulty in understanding the number 5,000,000 trees than I have in understanding the number 4,999,999 or 5,000,001.[257]

---

as Ross himself pointed out, the inability to see the difference leads to the problems we see in Feser's exposition of Ross's argument, namely the total disappearance of formal determination and of the rule-following considerations, and the emergence of conceptual issues that would otherwise not emerge.

[256]It is important to note that according to Aristotelians concepts, at least concepts of material substances like dogs, are the result of abstraction over particular dogs. Abstraction is the native human ability that allows us to hit upon and grasp what remains invariable throughout all the variables found in particular dogs (e.g. different colors, breeds, etc.). This is the Aristotelian approach to how the essence of things become the objects, or content, of thought (i.e. concepts). When we talk about the concepts of formal structures, however, there is the problem that if Ross's argument is correct these structures cannot be determinately abstracted from material structures—or, at least, not abstracted *in the same sense* in which we can be said to abstract the structure or essence (e.g., dogness) of material substances. Because material substances are formally indeterminate (in our sense of 'form'), any physical object or process that can be taken as an instantiation of a formal object or operation can only be an imperfect realization or instantiation of that object or operation. It will satisfy incompossibles, the way according to Feser a circular object in the material world will be indeterminate between a circle (and a qu-circle, to use a Kripke like example), a chiliagon, or a myriagon. But not so for non-formal material structures, since a dog is not indeterminate with respect to dogness and neither are human beings indeterminate with respect to the nature or essence of human beings.

[257] Celestine Bittle, *The Science of Correct Thinking: Logic*, rev. ed. (Milwaukee: Bruce Publishing Company, 1951), 24–8, quoted in (Feser, 2013, p. 8). Of course, the 5 million trees example works well with phantasms, but not with actual trees, which are even more obviously wholly material. For suppose there are 5 million trees in the Amazon, how would this be indeterminate, vague, or indistinct with respect to those trees amounting to 5 million trees? Our mental image would, of course, be vague and indistinct and would not differentiate 5,000,000 trees and 4,999,999, but not the actual trees in the Amazon. It makes no sense to say that the quantity of trees in the Amazon is vague and indeterminate except in the case of not having clear criteria for what counts as a tree (but then this would be equally a problem for the determining the precise content of the concept of 5 million *trees*). The fact that the quantity of trees in the Amazon is not indeterminate, that is, is neither vague nor indistinct, does not of course make the Amazon, or the trees in it, immaterial. Descartes' geometrical example is a much better one for illustrating an asymmetry between but the conceptual and the material *as such* as opposed to just the conceptual and the phenomenal or representational. Bittle's example is only good

Similarly, consider our concepts of a chiliagon (a polygon of 1,000 sides) and a myriagon (a polygon of 10,000) sides. Now let's consider their corresponding forms of representation. A chiliagon and a myriagon[258] are physically (that is, when represented physically as in a drawing, picture, or a computer monitor) and phenomenally (that is, when represented by mental images[259]) indistinguishable from one another, and they are in turn physically and phenomenally indistinguishable from a circle. But, obviously, these are different geometrical figures (we can prove this in geometry, of course), as our concepts of them show. But our physical and phenomenal representations of them (e.g. our mental images) are ambiguous and indistinct with respect to such figures. Not so for our concepts of them, which are clearly distinguishable from one another. Hence, our concepts of geometrical figures have a clarity and distinctness of content that their phenomenal and physical representations do not. Feser calls this feature of physical and phenomenal representations *semantic indeterminacy*. Furthermore, the physical and phenomenal representations of such geometric figures have features that the figures themselves do not and cannot have. Inversely, the geometric figures have features no material thing can have (e.g., no width).

> A Concept is equally representative of all objects of the same character. Thus if I see a circle drawn on a black-board, the concept which I form of that geometrical figure will express not merely the individual circle before me, but all circles. The figure I see is of a definite size, and is in a particular place. But my mind by an act of abstraction omits these individual characteristics, and forms the concept of a circle as it is enunciated in Euclid's definition. This concept is applicable to every circle that ever was drawn. When however I form the phantasm of a circle, my phantasm must necessarily represent a figure of particular dimensions. In other words the concept of the circle is universal: the phantasm is singular. Similarly, if I form a concept of 'man,' my concept is applicable to all men. But a phantasm of a man must represent him as possessed of a certain height, with certain features, with hair of a definite colour, etc.[260]

---

for illustrating an asymmetry between concepts and their corresponding phantasms, but not between the conceptual and the material *as such*. To be fair, these examples were introduced by Feser to differentiate intellectual from other mental phenomena.

[258] That is, their representations.

[259] Assuming that our mental images in question purport to represent how the figures look like. If the mode of representation is linguistic (say, like visualizing or hearing the word "chiliagon"), then what can be said about this form of representation is that the connection between the word "chiliagon" as a form of representation and the chiliagon as the geometric figure represented is arbitrary and conventional, without resemblance unlike the visual representation or image of a circle. Feser gives the following example: the image of a circle resembles the concept of a circle in a way the word "circle" does not, much like the image of a man resembles the concept of man in a way that the word "man" does not. (If you find this odd, it is. If you drop "concept of" in the example, the oddity sort of vanishes: the image of a circle resembles a circle in a way the word "circle" does not, much like the image of a man resembles a man in a way that the word "man" does not. It is hard to see how an image of a man resembles the *concept of* man if, as Scholastic-Aristotelians holds, the form of man is held without matter in the intentional order, that is, in the intellect. The image of a "man" resembles equally well the image of a "quman". The image of man is conceptually indeterminate. What about man itself? Is the animal we call "man" indeterminate amongst incompossible concepts, e.g. "man" and "quman"?) Feser also mentions those very clear ideas of which no *reasonable* phantasm or representation can be formed, e.g. law, God, justice, inference, wisdom, and so forth, though some sort of phantasm always accompanies such concepts.

[260] George Hayward Joyce, Principles of Logic, 3rd ed. (London: Longmans, Green, and Co., 1949), 16–7, as quoted in (Feser, 2013, p. 7). As Henning Tegtmeyer well points out, "this claim is odd. Mere images are hardly ever that definite. Being definite holds for real particular men, but not necessarily for man images." (personal correspondence) I agree. But this is a problem for Feser, who espouses this view, not for us who reject Feser's version of Ross's argument. In a sense, it is not entirely inappropriate to say that Feser has hijacked and straight-jacketed Ross's argument to fit some sort of

Feser, following some Scholastic commentators, goes on to say that the asymmetry between the means of representation for concepts and the concepts themselves illustrates why concepts cannot be identified nor reduced to their means of representation, such as mental images, linguistic symbols, material objects, etc. Another question is how these considerations show that concepts (or conceptual content) are *not* material. Here's where Feser employs Ross's argument and the skeptical paradox. No physical thing on account of its physical properties alone can *mean* a pure function like addition. For any material thing (e.g., the physical sign '+') that is taken to mean determinately addition can be equally well be taken to mean quaddition instead or any other incompossible function. Thus, the content of formal concepts is determinate in a way no physical thing can be, and this includes symbols and signs which are physical items assigned some sort of meaning by convention.[261]

6. MISUNDERSTANDING ROSS: THE FESER-DILLARD DEBATE

Feser's formulation of Ross's argument is, in my opinion, problematic in several ways and is handicapped by his failure to recognize that, if anything, it is a corollary to the argument from *formal* determination. It is therefore easily subject to certain objections (see chapter 5, §6, §7, and §8) that would otherwise dissolve if the main argument were primarily considered from the formal rather than the semantic standpoint. Rule-following considerations are simply lost on Feser's version of the argument, where the crucial point is the determinacy of what the thinking is *about* when it is about a formal rule or pure function. Thus, Feser's response to Dillard's epistemic objection (e.g., mitosis/schmitosis) is unsuccessful (see next chapter §8), for the same reasons that such a response failed and would fail against Goodman's green/grue examples. Feser (Feser, 2017, cf. 2013) attributes the following premise to Ross:

(P) All formal thinking is semantically determinate.

I have the following two general objections:

(1) To attribute proposition (P) to Ross is commit him to having espoused a false proposition (which would therefore make his argument unsound), and

(2) At any rate, Ross did not hold to (P), nor does his argument depend on (P) in any crucial way.

Briefly, the reason for (1) is that (P), whether as a statement of fact or as statement of an essential feature of formal thinking, seems to be downright false, for it makes thinking in purely formal terms

---

neo-Thomistic narrative *which Ross does not fit into*. In doing so Feser distorts Ross's original argument. This explains, in part, why Feser misunderstands Ross's argument and blends it with statements from a host of Neo-Thomistic authors that have very little to do with Ross's main line of reasoning. Dillard is right in objecting to Feser on these gounds as well. Now, my hunge is that the misunderstanding is due to the fact that in Neo-Thomist thought, the notion of 'form' when predicated of the understanding is used to mean something like conceptual content, viz., that which the thought is about is called an intentional form or a form as it exists *in ratio*. I am no expert on this, but from what I can tell, it seems that the confusion stems from reading Ross's argument through this peculiar neo-Thomistic lens.

[261] Note that Dretske (Dretske, 1986, 1981) has a broader notion of 'meaning', which encompasses natural meaning, functional meaning, and natural functional meaning. Feser is strictly concerned with the notion of 'meaning' as synonymous with concepts or conceptual content.

impossible and thus would not be able to account for our ability to construct pure formal systems such as context-free grammars or formal languages with what Hilbert called an ideal or infinitary sector, for example, semantically vacuous-indeterminate languages or systems in which formally determinate inferential or syntactical rules and operations can be carried out. Even in algebra we abstract away in many cases from content in our reasoning and in our formulation of formal rules (like a purely formal characterization of a logarithm or of the rule of transposition). Since we obviously can and do such things, (P) cannot be true. I guess the question is, after all, whether there is any such thing as pure formal thinking. I'd day that formal thinking, as activity, can be semantically determinate but it need not be (in such a case, one would have semantically indeterminate placeholders in place of the variable component, e.g. uninterpreted symbols).

But even if, for the sake of the argument, (P) were true, it is simply not a proposition Ross held—at least for the purposes of the main argument. This is our claim in (2). (P) makes Ross's argument one about semantic (read: content) determinacy, as opposed to one about formal determinacy. Indeed, given (P), it is hard to see what the fuss about the 'form of the thinking' was all about in Ross's writings. After all, both formal and non-formal thinking are semantically determinate, and since, as Feser makes clear with his explanation of (P), what is crucial here is semantic determinacy of thought, it seems that formal features *qua* formal ultimately drop out as irrelevant for the purposes of the argument. Non-formal thinking would do just as well. If Feser responds by saying, "but there is something unique and special about thinking about *pure formal* rules or *pure* functions", I'd agree; but I'd say that what's special *about* it is the fact that in order to think about such forms one has to be able to think IN (terms of) those forms, or in Wittgensteinian jargon, in order to think about pure formal rules one has to be (the sort of being that is) able to follow those rules. Hence the difference between merely thinking about addition and actually *adding*.

Now, what Ross did hold was that one could give a parallel argument as to content. As he rhetorically asks, "Can the reasoning above be applied to content as well as to form? Not so neatly, but still, I think, substantially yes…" (Ross, 2008, p. 123), after which he went on to argue in the affirmative for the non-physicality of content. But his main argument concerned the *form*, not the content, of logical and mathematical thinking (an analysis he thought could be extended in terms of "form" beyond mathematical and logical thinking to all kinds of intellectual acts, such as asserting, denying, intending, doubting, etc.). But Feser's version of Ross's argument seems to ignore or, even worse, obliterate that difference and is thereby threatened by conceptual confusion—e.g., such as the conflation of *thinking about addition* with *adding* or thinking about *modus ponens* with applying and carrying out *modus ponens*. This is evident in the following passage,

> Now, Ross argues, first, that at least some of our thoughts and thought processes do have a *content* that is entirely determinate or exact. The sort of thinking involved in mathematics and formal logic is the example on which he focuses. There can, for instance, be an objective fact of the matter that I am *adding*, specifically, or reasoning according to the inference rule *modus ponens*.
>
> (Feser, 2017, cf. 2013)

Notice the shift from the semantic aspect of understanding (i.e. "a content that is entirely determinate") to the formal aspect of the understanding or act (i.e. "an objective fact of the matter that I am *adding*, specifically, or reasoning according to the inference rule *modus ponens*"). Examples like this one abound in Feser's treatment of Ross's argument. There is a constant conflation between the form as content and the form as form. Whatever it is that makes it an objective fact of the matter that I am *adding*, this should be cashed out in terms of the form of the thinking not the content of the thinking. Note that since logical and mathematical operations will ever be carried over

a finite subset of the set of possible variable components that such operations can have, the form of the operation cannot be determinately sorted out by any *specific* content (that's one of Ross's great insights, e.g., the content of quadding and adding, for any pair of numbers less than 57, is exactly the same); so, no semantic considerations will ever enable us to capture and settle definitively the kind of formal activity one is or has engaged in, since that is determined by the form (the invariable component) of the thinking not by the content (the variable component). So, it seems to me, that by representing Ross's argument from the formal determinacy of understanding as being instead an argument from semantic determinacy, the truly insightful, original, and revolutionary aspects of Ross's argument are lost. It also shows that Feser is not clear about the difference between the form and the content of reasoning, being guilty of running them together as conceptually indistinct.

Furthermore, there are some unique considerations to the parallel semantic version that Ross himself develops which do not figure at all in the version of the argument that Feser countenances.[262] This is not the place to consider them in detail here. But here are two of the reasons (cf. Ross, 1992, p. 150, n. 17) provided by Ross for why an argument analogous to the one from formal determination might, substantially, be applicable to content (Ross, 2008, p. 123).

(i) thought content has to be repeatable and shareable in ways the physical cannot be, for instance, shared by another person;
(ii) thoughts lack the transcendent determinacy of the physical.

Feser seldom addresses these two crucial points. Both (i) and (ii) have formidable considerations behind them, but an in-depth analysis of them would involve a rather long detour into metaphysics, philosophy of nature, the nature of abstraction, philosophy of mathematics, and the philosophy of science. Ross develops these considerations in detail in published articles as well as in his last book (see Ross, 2008). But let me briefly say something about (ii). The transcendent determinacy points to the fact that the compliant reality that constitutes the truth-maker of our judgments about the material or physical world includes features, necessities, and processes that go well beyond anything contained, and that could be contained, in the judgment. It is not merely that a host of different realities can make the statement 'Someone is knocking at the door' true, but that in a given case the compliant reality will make that statement true on account of features that do not make it into the judgment but that nevertheless are part of the truth-makers of the judgment. Thoughts, judgments, and understanding in general lack such a transcendent determinacy, Ross argues. What makes them true is always more than what is contained or could be contained in/by them. The compliant reality (the world, nature) 'overflows' any judgment that could be made about it, though this is not the case when the compliant reality are formal realities (hence, Ross's rejection of Platonism about formal truth).

In fairness to Feser, Ross himself talks about pure functions as the structure or form of the thinking and also as the content of thinking. For example, "… each [understanding and the bodily medium in-which it exists] has features the other cannot have (determinate, repeatable content for judgments versus transcendent, unrepeatable determinacy for the bodily states." (Ross, 2008, 125) But this corresponds to our conceptual distinction between the formal aspects and the semantic aspects of understanding (see chapter 1), and that once one has established the insufficiency of any material principle for determining the formal aspects of the understanding, matter will be insufficient to determine the content of the understanding when the content is a potential formal

[262] Some, however, like universality for example, do play a role in Ross's argument but only a minute one and of a somewhat different nature to Feser's (see e.g., Ross, 2008, p. 125-126, for the discussion of universality, the explicit statement that it applies to *both* form and content, and that it concerns an ability).

structure or aspect of the understanding, i.e. when the content is *formal* content (e.g., thinking about addition). So, the primary focus of Ross's argument lies in the *forms* of the understanding, not in content *qua* content. Yet, as has been remarked, Ross considers a further asymmetry between the cogitative and the material on the basis of the transcendent detrminacy of the physical and the lack of such transcendent determinacy of the content of the understanding.

> Thought, as content, is immaterial in other ways. For instance, it lacks the transcendent determinacy of the physical. A true judgment "someone is knocking on my door" requires for its physical compliant reality a situation with many features not contained (or logically implied) in the true judgment: thus, various determinate but incompossible physical situations might make the same statement true.
>
> (Ross, 2008, p. 199 n. 24)[263]

There are two important implications that follow from this argument from content and our argument from form. First, that pure formal structures (e.g., formal truths, rules, concepts, functions, and operations) lack transcendent determinacy. This also means that in the case of formal judgements, judgments about pure formal structures, the compliant reality is somehow logically contained and implied in the judgement *in toto*.[264] That is, formal reality or formal being lacks the transcendent determinacy of the physical: the being of formal structures is consequent upon human productive and creative understanding, whereas the being of material structures is antecedent to human understanding of any sort.[265] So, the compliant reality of formal judgments (e.g., numbers in the case of judgments about numbers) cannot be physical. There are no determinate but incompossible compliant formal realities that make formal judgments true, though physically determinate but incompossible physical realities might satisfy such judgments derivatively. The original or non-derivative compliant reality of formal understanding and judgments must be formal, not material (hence, e.g., mathematics cannot be a branch of physics). Secondly, if formal and non-formal understanding lack the transcendent determinacy of the physical, then the physical has features that understanding cannot have. Hence, it provides an alternative way of proving the ontological asymmetry between the understanding and the material on the basis of a feature that material reality has that understanding cannot have. In a way, there *is* something special about matter.

Be all that as it may, this is not the place to canvass the considerations behind, and the merits of, the parallel argument as to content. Suffice it to make the following two points. First, that the semantic aspect of understanding may point to a further ontological asymmetry between the cogitative and the material, but the asymmetry would obtain in virtue of features that are entirely

---

[263] "Further, some thinking is as much physical as it is immaterial. My walking, as an action, is as much a mode of thought as it is a movement, yet no movement, however complex, could ever make a thought. Leibniz, in section 17 of the *Monadology* … says that if perception were supposed to be produced by a machine, we could make the machine on large scale and walk around in it like a mill; we would never find a perception, only movements of wheels, gears, and pulleys. […] Further still, machines do not process numbers (though we do); they process representations of them (signals). Since addition is a process applicable [directly, non-relatively] only to numbers, [physical] machines do not [directly, non-relatively] add. And so on for statements, musical themes, novels, plays, and arguments. Machines process only representations, but the pure functions are among the represented." (Ross, 2008, pp. 199-200 n. 26)

[264] For a more robust explanation of what this claim amounts to, see (Ross, 2008, chapters 1, 2, 4, and 7 in particular). For more on Ross's view concerning the nature on pure formal truth and structure, see (Ross, 2012, 1992, 1990, 1989, 1988, 1986)

[265] This is admitedly anti-Platonic and anti-Augustinian (and, therefore, anti-Thomistic). There are no formal truths in the absence of human intellects. See n. 264 for references to Ross's work on the nature of formal concepts and formal truth. Needless to say, KRA need not be committed to this view of formal structures.

different from those considered by Feser in his semantic version of Ross's argument. Secondly, the argument from *form* stands or falls independently of Ross's argument from *content*, which is perhaps better called "Ross's argument from the transcendent determinacy of the physical." (For a fuller defense of the transcendent determinacy of the physical, and the difference between physical truth and formal truth, see (Ross, 2008, pp. 11–84, 129–149).) The argument from the transcendent determinacy of the physical is, however interesting, not a corollary of the argument from form and our goal here has been to develop and refine the argument from form, not the other one. The merits (or lack thereof) of the argument from the transcendent determinacy of the physical is a task left for another occasion.

7. CONCLUSION

In this chapter we have argued that if Kripke's argument for the formal indeterminacy is sound and if it is the case that formal understanding is physical, then formal understanding would be formally indeterminate, in which case there would be no such thing as formal understanding and activity (e.g., no such thing as doing mathematics, or having the ability to do mathematics, in any objective sense). Take logic. Let *R* be a formal rule, namely *modus ponens*. Let *R** be a qu-like rule incompossible with *R*, say *modus quonens*. Then, for any φ (let φ be any physical mechanism, process, etc.), if φ is said to realize *R* , then φ can be said to realize *R**, and conversely. We have argued in a variety of different ways that this cannot be true of formal understanding, so 'formal understanding' cannot be part of the domain or universe of discourse of φ which is the domain containing all material things. Understanding must intrinsically, determinately, and objectively contain or realize pure functions, otherwise we couldn't do what we do such as add, conjoin, square, reason validly, and so on. So formal understanding cannot be physical, for it is formally determinate (alternatively, the capacity for formal understanding cannot be physical because it is formally determinable). Plus, claiming that the understanding is formally indeterminate consists in a performative contradiction, leads to absurdity (it is self-refuting), and is, ultimately, wholly unintelligible. Therefore, formal operations cannot be wholly a matter of matter, viz., "there are no logical relations in [matter], any more than there is genuine addition as leaves pile up" (Ross, 2008, p. 198 n. 16) and "… the judgment 2 + 7 = 9 [or the judgment if *p* then *q*, *p*; therefore *q*] cannot be identical with a physical state or process that is theoretically indeterminate as to its structure, as all physical states are." (Ross, 2008, p. 124)

We have considered some objections, but many more remain. We now turn to some of these.

5
Objections and Replies

In this chapter we consider some objections to Ross's argument in particular and to KRA in general. Since the argument is not well known in the literature, some of the objections presented here were originally raised in entirely different contexts (e.g. model theory) and for entirely different reasons, yet we've adopted and adapted them *mutatis mutandi* to our present discussion and argument (KRA and the nature of formal understanding). The choice has not been arbitrary. These objections have been included because we've deemed them (or some variant thereof) as potential objections to *our* argument. On the other hand, given that KRA is in general unknown to the philosophical community, some of the objections we will consider are of our own making but we have tried our best to come up with the most compelling and obvious objections to KRA. We will argue that none of them conclusively succeed in refuting the formal determinacy of the understanding. In the last three objections, quite a bit of attention is payed to Edward Feser's (Feser, 2016, 2013, n.d.) construal and appropriation of Ross's argument; for he is (as far as I know) the only one to have discussed at length and appropriated in published work Ross's own argument. The last three objections (6, 7, and 8 below) are objections that can be raised against Feser's version of Ross's argument but which, we argue, cannot be raised against KRA.

1. THE EPISTEMIC OBJECTION

Our discussion of KRA up to this point has had much to say about epistemological qualms. The reader should be familiar with what I call 'the epistemic objection' to Kripke, to Ross, and therefore to KRA. We'll state it as follows:

> **The Epistemic Objection (EP)**. Kripke fallaciously inferred an ontological conclusion (i.e. formal indeterminacy) from epistemological findings (i.e. formal underdetermination). From the fact that it cannot be determined, not even on the basis of an exhaustive and maximally specific physical description of a physical system, which among several incompossible pure formal structures the physical system realizes it does not follow that it doesn't intrinsically realize any. All we can say is that it cannot be *known* or *established* which formal structure a physical system realizes and that therefore the physical is at most formally underdetermined. Such findings concern our epistemic limitations and warrant no claims about any intrinsic (metaphysical or ontological) limitations on the part of the physical. Therefore, all the Kripkean argument shows is that the physical is formally underdetermined, not that it is formally indeterminate. Therefore, Ross is not warranted in inferring an ontological asymmetry between the physical and the cogitative. At most, all we can establish is that there's an epistemic asymmetry.

This is an objection that comes up rather frequently Kripke's argument or KRA more generally is brought up in some discussion. There are several ways to address this objection. Dialectically, we can counter to the epistemic objector with the following response. Even if we take the bull by its horns and accept the premises of the objection, the epistemological conclusion would still be devastating for any type or token physicalism about the realizers of pure formal structures, since no such formal underdetermination, which holds in regard to any purported physical realizer, holds when the realizer under consideration is human intellectual activity.

Here's why human intellectual activity is not and cannot be formally underdetermined (in the epistemic sense). Formal underdetermination of our intellectual powers and activity would entail that we never know what we are doing when we are doing algebra, calculus, number theory, set theory, modal logic, and so forth, and thus we would not know that, or what, we are proving whenever we carry out formal proofs (e.g. it would've been epistemically indeterminate to Gödel whether he was proving, or had proved, his Incompleteness Theorems—and to any of us for that matter whenever we try out the proofs ourselves). Furthermore, to claim that our intellectual activity is formally underdetermined would make it the case that we cannot even know whether we are *arguing*, let alone arguing for the formal indeterminacy or for the formal underdetermination of the physical since we cannot even know whether we are arguing at all! An argument is constitutively of a determinate logical form, whether a valid one or an invalid one; but since it is *ex hypothesi* formally underdetermined which logical forms are being realized whenever we reason, we cannot know of which logical form our reasoning is a case and thus cannot even know whether we are arguing. But, of course, we *do* know that our understanding is determinate with respect to such forms, as was shown above. Hence, it cannot be the case that our intellectual powers are formally underdetermined. But we have already shown (and the objection grants) that the physical is *at least* formally underdetermined.

Now consider the following argument:

(1) If intellectual activity is metaphysically formally indeterminate, then it is epistemically underdetermined that intellectual activity is metaphysically formally indeterminate.[266]
(2) Intellectual activity is not epistemically formally underdetermined.
(3) Therefore, intellectual activity is not metaphysically formally indeterminate.

If one holds, as Dennett would, the ontological or metaphysical position that human thinking is formally indeterminate, and adds the premise that whatever is formally indeterminate is also formally underdetermined, the human understanding cannot be formally indeterminate since it is not formally underdetermined. But (1) is contentious. Things can be metaphysical indeterminate and yet known to be so. Indeed, *pace* Kant, that is our point. Furthermore, with regards to the forms of understanding, our position is even more forceful. For we do carry out determinate formal operations and do grasp pure formal rules or functions. The very exercise of formulating and comprehending Kripke's argument is an instance of formal understanding at work. We know and understand a variety of mathematical and logical functions, we can compute many such functions and apply formal rules, we do carry out formal operations, and so on. So our understanding is not formally underdetermined. But if this argument were sound it would remain that human intellectual activity is, in both the epistemic and metaphysical sense, determinate in form and thus fundamentally different from anything physical. In does not matter if we modify the argument so as to be sensitive to the fact that not everything that is metaphysically indeterminate in form is also epistemically underdetermined with respect to form. A modified version of the argument that takes this into account would look like this:

---

[266] Talk of 'epistemic underdetermination' seems a bit redundant since 'underdetermination' is by definition epistemic or an epistemological notion. But it is my experience that this is easily confused or forgotten, especially when one formulates an argument that using both 'determination' (which can be used either in a metaphysical or an epistemic sense) and 'underdetermination' (which is understood epistemologically). This might explain why the epistemic objector thinks Kripke and Ross are taken about underdetermination issues when the say that physical properties alone do not suffice to determine a particular pure formal structure. With respect to the former notion of 'determination', unless I specify otherwise (as in e.g. "epistemically determined" or "epistemic determination"), the default use of the term will be ontological and metaphysical.

(1´) If intellectual activity is metaphysically formally indeterminate, then either it is epistemically underdetermined that intellectual activity formally indeterminate or it is epistemically determinate that intellectual activity formally indeterminate.
(2´) It is epistemically determinate that intellectual activity *is* formally determinate.
(3´) But to say that it is epistemically determinate that intellectual activity is formally determinate is to say that it is not the case either it is epistemically underdetermined that intellectual activity formally indeterminate or it is epistemically determinate that intellectual activity formally indeterminate.
(4´) Therefore, it is not the case that either it is epistemically underdetermined that intellectual activity formally indeterminate or that it is epistemically determinate that intellectual activity formally indeterminate.
(5´) Therefore, it is not the case that intellectual activity is metaphysically formally indeterminate.

The consequent of (1´) is a disjunction whose disjuncts turn out to be false. So we can negate the disjunction, which is equivalent, by De Morgan's theorems, to the conjunction of the each negated disjunct. By *modus tollens* we get the negation of the antecedent of that conditional (1´). Therefore, it is not the case that intellectual activity is metaphysically formally indeterminate. Hence, human intellectual activity *is* known to be formally determinate and there *is* formally determinate. We showed this much in chapter 4 *contra* Dennett.

To say of something that it is underdetermined is to say that we cannot decide (i.e. come to know) which of the available competing but logically incompossible structures (descriptions, hypotheses, theories, models, functions, etc.) is true or false of a given thing. In such cases we are unable to epistemically determine which (if any!) of the incompossible pure forms the physical at any given moment satisfies, implements, or realizes. To say, however, of something that it is epistemically determinate (or determinable)[267] with respect to a model or a theory is to say either that it is known (or knowable) that it realizes such a model or theory, or that it is known (or knowable) that it realizes *no* such model or theory. In other words, to say of something that it is epistemically determinate with respect to ψ is to effectively deny underdetermination of that thing with respect to ψ, for it is to assert that we *do* know whether or not it is metaphysically determinate with respect to ψ. Since we do know that human understanding is determinate in form, then human understanding *is* determinate in form. We cannot know any such thing about the physical.

Unfortunately for the materialist, the epistemic objection does not do her any service. She would have to show that it is epistemically determinate that we do *not* think, judge, or reason in terms of such forms or formal rules, and thus that we cannot understand them. But there is no showing this without understanding them, without a grasp as to what is being denied or without at least performing an act determinate in logical and grammatical form such as negation (e.g. "it is not the case that we think in logical forms"). This is also why arguing for the formal underdetermination of understanding does not work. Thus, there's something true about intellectual activity that cannot be true about the physical: that the former, unlike the latter, is *not* formally underdetermined. For we can know that the cogitative is and can be determinate in form, but we do not and cannot know such a thing about the physical. This still seems to imply an 'in principle' difference between the two. In the words of Ross,

> Distinguish reasoning that the form of thought is definite in the way the physical cannot be (the main line of argument) from the epistemic version that we can know

[267] For 'epistemically determinate' read 'it is not underdetermined'.

> that thought is determinate in form though we cannot ever know that a physical process is similarly determinate. *The latter would also imply that thought and the physical are different in principle*. But it hasn't the same weight as the reading I rely upon.
>
> (Ross, 2008, p. 196 n. 4; my emphasis)

It hasn't the same weight because, whereas the ontological and metaphysical version decisively establishes that the physical cannot be determinate in form, the epistemic version —according to the objection—does not establish such a claim, leaving it open whether the physical *could* be determinate in form (though not leaving it open whether it can be known to be so; it effectively establishes that it cannot). The epistemic objection, however, makes things such that the onus is on the token or type physicalist to give an account of how the physical could determinately be of a given pure form, even though such a thing could never be known, as well as to explain how such an outstanding epistemic asymmetry is possible without there being an ontological and metaphysical difference. But notice that since the token physicalist, whether functionalist or not, accounts for mathematical and logical reasoning in terms of the physical realization or exemplification of mathematical and logical forms, and such realization or exemplification is underdetermined (epistemically indeterminate) and thus inaccessible, then what we do when we reason logically and mathematical should be underdetermined and inaccessible.[268] But it is isn't. By the very fact that it is *not* underdetermined or epistemically inaccessible, we effectively show that intellectual activity cannot be wholly physical.

The epistemic version can also establish a difference in principle between the physical and formal understanding in another way: namely, by bringing to the fore in Wittgensteinian fashion that whatever formal determination the physical can be said to have, it cannot be the same as the understanding is said to have, otherwise it would be

(a) in principle possible to know that the physical *could be* determinate in form, and
(b) in principle possible to say truly and intelligibly the same things about both, such as "it is self-refuting and incoherent to affirm that x is formally indeterminate".

With respect to (b), we know that 'formal indeterminacy' cannot intelligibly hold with respect to human understanding but we *do* know that it can (and does) intelligibly hold with respect to the physical. A condition under which the concept 'being formally determinate' would apply to x is if it were absurd and self-contradictory to deny the concept to x. But it is not, as the Kripkean argument shows, if the domain of x is the physical. Yet it is if x ranges over the cogitative.[269] So we have an asymmetry in intelligibility conditions between the physical and the cogitative. With respect to (a), to be able to talk about the physical being 'formally determinate' or 'formally determinable' we must be able to assess the conditions of applicability of that concept, that is, suitable conditions under which the concept could be intelligibly applied to x so as to have conditions under which it could be true (or false) of x**.** But we *do* know the conditions of applicability of that concept with respect to the

---

[268] Note that this makes the functionalist or computationalist accounts of the cogitative that buy into some version of physicalism subject to the implication that, with respect to logical and mathematical concepts, rules, and operations, the cogitative is underdetermined. So we could not know whether we were reasoning validly or invalidly, for example, for we could not know whether we were reasoning logically at all. Likewise, there'd be no knowing whether one is, or even can do, mathematics since any activity will be mathematically underdetermined. This is, at any rate, self-refuting and would be unintelligible since there would be no saying what mathematical or logical underdetermination is given that any such saying would be in turn mathematically or logically underdetermined. We wouldn't even be able to understand what it is for something to be 'mathematically underdetermined', since these requires a determinate grasp of the mathematical theories, specifications, or models which something underdetermines.

[269] If it is not clear by now, I use 'cogitative' or 'cogitative powers' as a shorth-hand for our formal capacities, abilities, and activities. So 'the cogitative' encapsulates the capacity for formal understanding, formal understanding as such (as actualized capacity in acquired abilities with respect to formal concepts and rules), and formal activity (as the actualization or exercise of the relevant acquired abilities).

cogitative and we *do* know that neither physical nor phenomenal facts and features satisfy them, for theirs' are *observer-dependent conditions of applicability*; so if 'formal determinacy' is said to apply to the physical nonetheless, as we surely say it can, it must do so under different conditions of intelligibility than those under which it applies to the cogitative and therefore we must be dealing with different, either equivocal or analogical, senses of 'formal determination' when predicating formal determination of a physical mechanism on account of its physical properties alone and when predicating formal determination of the cogitative. But given that the whole point of the epistemic objection is to point out that since the problem of whether the physical is formally (in)determinate is (or might be) epistemically inaccessible, unknowable, one cannot infer indeterminacy from underdetermination, we are left with the mystery of how the concept of formal determination can be *known* to be intelligibly applicable to and true of the cogitative but not the physical. The use of the concept of 'formal determination' in the epistemic objection, where it is used to apply to the physical, is left as utterly mysterious, being epistemically inaccessible the way (if any!) in which it could be intelligible and true that the physical is something formally determinate (and thus, how the relevant formal concepts could intelligibly and truly apply to anything physical). But the epistemic objection concedes that the conditions of applicability (intelligibility conditions) are not the same between the physical and the cogitative. So the concepts cannot be the same; in other words, formal concepts cannot apply in the same sense to the physical and to the cogitative. In chapter 3, we explained though only briefly one way in which the concept of computation, one amongst many formal concepts, could intelligibly and truthfully apply to physical mechanisms. But it hinges on the recognition that it not *univocal* with the way in which 'computation' is truthfully and intelligibly said of a competent person's application of formal rules for the computation of a given mathematical or logical function. The point being the following: the expression 'being determinate or determinable in (e.g. mathematical or logical) form' as it applies to the cogitative cannot be univocal with that expression as it applies or can apply to the physical, and the epistemic objection actually *concedes* this much. So we *do* know that even if the physical is in some other sense formally determinate and determinable (in what sense?) it cannot be univocal with the formal determinacy and determinability of the cogitative. Hence, human understanding and the physical are still *in principle* different.[270] Let me note that we can ask the epistemic objector something that needs to be addressed before a suitable answer to his objection can even be given: what would it take to *know* that the physical is formally indeterminate *in se*? That is, what would count as *not* knowing that fact? Unless this is answered, it is unclear that our argument has not shown that we *do know* that the physical is formally indeterminate.

2. THE OBJECTION FROM NON-STANDARD MODELS: SKOLEM AND KRA [271]

The Objection from Non-Standard Models (the Löwenheim-Skolem Theorem, LST). Formal understanding is indeterminate, because any formal reasoning in first-order logic has non-standard models. That is, formal reasoning cannot distinguish the intended model from the unintended model (non-standard model) and so is indeterminate. The objection only applies to formal reasoning--in a formal language. So formal understanding is subject to 'Skolem's paradox': The same set of

[270] If the epistemic objector is still unconvinced about the inference from conditions of applicability (intelligibility) of a concept to there being an intrinsic and in principle difference between the things that satisfy and those that do not (and cannot) satisfy such conditions, she can perhaps take the inference to be abductive, rather than deductive, namely an inference to the best explanation. The onus is on her to offer a non-*ad hoc* explanation of the asymmetry in intelligibility conditions between the physical and the cogitative *vis-à-vis* formal concepts in a way that does not involve the nature of the physical and the nature of the cogitative.

[271] I thank Prof. Buechner for raising this objection.

sentences in the language used to reason about the real numbers (which are nondenumerable) is also satisfied by denumerable sets—such as the natural numbers.

Notice that in its philosophical significance and conclusion the objection is, ultimately, a variant of the Q-D objection: human understanding, at least with respect to first-order logics, cannot be formally determinate. It is admittedly quite different from the Q-D objection insofar as it uses a fundamental result in the model theory of first-order logics to undermine KRA (for a useful discussion, see Bays, 2014; Hodges and Scanlon, 2018; cf. George, 1985 for a discussion of the philosophical significance of the Löwenheim-Skolem theorem). Note that the scope of this objection is limited in ways in which the Q-D objection is not, namely to reasoning in first-order logic. But KRA applies to much more, including second-order logic and our ability to do second-order logic. But let us address the objection since some philosophers (i.e., Quine) have championed the idea that logic *is* first-order logic. The theorem is quite technical; fortunately, we do not need to go into the technical details of the theorem in order to answer the objection.[272] A brief sketch will do for our purposes. Our response will be quite general and will mirror our response to the Q-D objection.

Very roughly, LST amounts to the following.[273] A model is an interpretation that assigns the value 'true' or '1' to that which it interprets. A model of a first-order formal language L is an interpretation of the sentences of L that makes all the theorems of L true. The idea then is that there is a standard model M and a non-standard model M^* which L satisfies (e.g., if it helps, think of L as a first-order formal language sufficiently strong to express an elementary amount of arithmetic whose grammar, which contains the symbol '+', satisfies a standard interpretation where '+' denotes the addition function and a non-standard interpretation where '+' denotes the quaddition function and that both interpretations make every theorem of L true. Hence, L cannot exclude 'incompossible' models from being equally true of it). To say that this is a crass oversimplification of LST would be an understatement. But it will suffice for our purposes. Suppose by 'formal understanding' and 'formal reasoning' we mean 'understanding first-order logic' and 'reasoning in accord with the rules of first-order logic'. Now consider what would have to be the case if the objection from LST were correct.

(1) Formal understanding and reasoning would satisfy standard and non-standard models and therefore would be indeterminate amongst incompossible models.
(2) But if formal understanding and reasoning are indeterminate amongst incompossible models, then there is no objective fact of the matter as to which formal rules, operations, functions formal understanding and reasoning are instances of or are in accord with since any instance that can be said to be a case of a standard model can be said to be case of a non-standard model.
(3) But then it follows there is no fact of the matter (no objective, interpreter-independent fact or truth) about whether we can understand what a standard model is, since this would entail knowing what it is to reason in accord with a standard model and thus exclude incompossibles, and what a non-standard model is, since this would entail understanding

[272] A paper is in the works that compares Gödelian arguments against computationalism to Kripke's argument against computationalism and in which the relevance and significance of the Löwenheim-Skolem theorem is considered at greater length and depth than here. A draft of this paper was presented at 'The Singularity Summit Imagination, Memory, Consciousness, Agency, Values' on September 26, 2018, in Jesus College, Cambridge.
[273] For a rather technical exposition of the general (upward and downward) Löwenhein-Skolem theorem, see e.g. (Mendelson, 2015, pp. 127–129). A more accessible exposition can be found in (Bridge, 1977).

what it is to reason in accord with a definite non-standard model and thus exclude incompossibles.

(4) But then no one could prove nor understand the proof that formal understanding and reasoning, that is, understanding and reasoning in accord with first-order logic, satisfy (countably many) non-standard models.

(5) Hence, no one could prove and understand LST.

(6) Hence, no one could intelligibly object to KRA on the basis of LST.

But,

(7) Since one *can* object to KRA on the basis of LST, one *can* come to understand LST, and one *can* prove LST (as Thoralf Skolem himself did), it follows that

(8) It is not the case that formal understanding and reasoning satisfy standard and non-standard models and therefore is indeterminate amongst incompossible models.

If the objection from LST were true, it would be wholly unintelligible to us. For it would be wholly unintelligible what a theory of first-order logic is, what a standard model of it is, and what a non-standard model of it is. A standard and a non-standard model of first-order logic, for example, will assign divergent but formally determinate functions and operations to the very same formal language of the logic. So first-order logics satisfy non-standard models, yet nothing of significance about formal understanding and reasoning as such follows from this. First-order logics can be made model-theoretically determinate *by choice* or *fiat*, of course; the logician decides. Intrinsically, they remain model-theoretically indeterminate, though syntactically determinate. First-order logic cannot decide nor choose which model is true of it, but we *can* choose a model for it (whether standard or non-standard) though strictly speaking the chosen model would not be 'true' of it in the sense of making any other applicable one objectively 'false'. But if the logician's understanding is formally indeterminate, there is no such thing as the logician understanding the distinction between standard and non-standard models and choosing, say, for heuristic purposes, one model over another. In fact, as anyone who has worked up to the problem of non-standard models in model theory, one must first begin by studying first-order logic under the standard model. If this is indeterminate, no such thing would be possible. In short, LST is a manifestation of formally determinate understanding and impossible to grasp if one could not determinately grasp standard and non-standard models, the differences between them, and the reason why any first-order logic will be subject to LST. At any rate, this objection leaves it unexplained how we can understand and do second order-logic (to which LST does not apply) and prove theorems (in model theory, for example) about first-order logics. If LST implied the formal indeterminacy of formal understanding, none of this would be possible.

3. THE PSYCHOLOGISM OBJECTION

This is a problem that was, albeit in a different guise, anticipated by Kripke in his brief discussion of Frege and realists ('Platonists') about mathematical entities (Kripke, 1982, pp. 53–54). Regardless of the metaphysical status of mathematical objects as such, it still remains that there is mathematical and logical activity. We do not merely grasp as extraneous content a given function or formal operation, we carry out formal operations and compute functions. Our cogitative acts are formally (mathematically or logically) determinate and such cogitative activity is *normatively* constrained by the nature of the formal rules or functions in question. On the basis of this observation, the anti-psychologistic minded philosopher about mathematics and logics might accuse us of reducing logical and mathematical operations to features of human psychology, thus making the seemingly non-mind dependent features of mathematical and logical structures a feature of the

human mind (Kusch, 2015; Pelletier et al., 2008; see the papers in Notturno, 1989). This objection is easily dealt with. First, in chapter 1 we acknowledged that our claim was not that formal rules and functions were reducible to mere forms of understanding but rather that they had to be at least forms of human understanding, otherwise there could be no cogitative act that is formally determinate and which therefore has a formal component or structure that makes the very act a real bearer of formal features (e.g., deductive validity). In other words, we could not do what we do when we do mathematics and logic, when we reason mathematically and logically. We also noted that these rules which constitute formal operations are normative and therefore stand as standards of correctness to their applications. Formal rules determine what it is to apply them (and thus, epiphenomenally perhaps, what it is to misapply them). There is no reduction of the normative to the descriptive, let alone a reduction of the purely formal to the neuro-chemical properties of the brain, to mere regularities found in intellectual behavior, or what have you (cf. Frege, 1980; Husserl, 1982).

In fact, our argument is perhaps anti-psychologistic *par excellence* since it denies that the determinacy of formal understanding is a consequence of, or reducible to, some sort of material mechanism (e.g., the human brain) or some sort of causal-dispositionalist, descriptive, regularity found in humans. What the objection may be pointing at is that we do deny that there is a strong gap between formal understanding and formal operations (formal rules, algorithms, functions, etc.). We have distinguished formal entities and objects from formal operations (see chapter 1) in order to make it clear that, even if formal entities and objects exist in the way in which the Platonist would want to assert (independent of any human mind), the same thing cannot be said with respect to formal operations (e.g., the host of formal operations and acts that fall under the umbrella terms of 'calculating', 'inferring', 'proving', and so forth.) since these are, minimally, things we (can) *do*. Think of them as structural principles for intellectual activity that order and structure the understanding and activity in the requisite ways, which in the present context are mathematical and logical. Maybe there's a Platonist argument for formal operations and a way to make sense of the fact that we do things with (carry out formal operations over) those Platonic formal entities, but it would have to respect the intrinsic and non-derived formality of the understanding and the formally determinate acts (operations) that obtain upon its exercise. This might make the human *psyche* special or *sui generis*, but it won't make our argument a psychologistic one—at least not in the negative sense in which 'psychologism' has been traditionally understood (Cohen, 1998; Mohanty, 1997; see again Frege, 1980; Husserl, 1982). But in the sense in which the formal structure as well as the semantic content of formal operations is "in some way" (Pelletier et al., 2008) a feature of human psychology—by which we mean e.g., 'is present in an inherent, intrinsic, and objective manner in the sort of cogitative acts that are the exercise of formal the understanding'—then our argument can be said to be psychologistic. But not in any naturalistically relevant and interesting way, contra (Pelletier et al., 2008). KRA implies that the aspects of human psychology that are responsible for logical and mathematical understanding cannot be naturalized and somehow transcend (though not entirely) the limitations and constraints imposed by its material (neurobiological) underlying necessary mechanism (cf. Chomsky's and Fodor's claims to the effect that our linguistic or cogitative abilities are in principle capable of generating infinitely many syntactic-representational units by recursive operations). Again, with respect to the capacity for formal understanding, to the possession of formal understanding as such (the ability), and to the formal activity that obtains as a consequence of exercising one's formal understanding, human psychology cannot be wholly matter of matter and cannot therefore be naturalized.

4. THE ABSENT POSITIVE FACT OBJECTION

What I call the 'absent positive fact objection' consists in objecting that, since we have yet to provide a *positive* set of facts that could determine which formal rule someone is following in a given case, that our argument is somehow defective. The objection insists that we have not answered satisfactorily the constraints of the skeptical paradox, that we have not offered a positive account of formal rule-following, human computation, formal understanding, or whatever, that could replace the naturalistic one and that, in the absence of a positive account of our own that could replace naturalism, we might very well be justified in ignoring the significance and implications of KRA. This is a reasonable response, since one does not dispense with a seemingly valuable and useful architecture without a suitable alternative at hand that might, architecturally speaking, supplant the former. But as an objection to our argument and dialectical strategy, this objection misses the point. True. We have not provided a positive account in the sense of giving an explanation as to how it is possible for us to follow and apply determinately, objectively, and intrinsically (on account of abilities *we* truly possess) formal rules. A positive account of rule following in mathematics, logic, and so forth, has been provided by some philosophers, including Wittgenstein (Wittgenstein, 2009, 1996; see e.g. Hacker and Baker, 1984, 2005, 2014; Wright, 1981, 1989, 2007). Admittedly, this is a task that we will eventually have to undertake in order to test the cogency of KRA holistically and to make it more palatable for those who find its conclusion somewhat bittersweet. But the purpose of our argument was to show that no positive account for formal understanding and activity in terms of physical properties alone can be given. We did not embark ourselves on the project of giving a positive answer to the question, how is it possible for the cogitative to be formally determinable and determinate? Rather, we embarked on the project of showing that no naturalistic positive account of the cogitative can be intelligibly and coherently sustained. An absence of a positive, replacement account is not a refutation of our refutation of a given positive account. We acknowledge, however, that a substitute positive account of the cogitative, particularly in ways which respect the psychosomatic unity that human beings are yet allows for the transcendence of the cogitative, must be formulated. Yet this was never the task of the present work and therefore it should not be held against us that have not provided a positive account that could replace the naturalistic account of the cogitative. Suffice, for our purposes, to show that the naturalistic positive account is untenable.

5. THE MEREOLOGICAL OBJECTION

This objection is inspired by a hypothetical Wittgensteinian objector, given similar objections raised by Wittgensteinian against explanations of activities that ascribe to parts, components or subcomponents, activities that can only be properly and intelligibly ascribed to the whole (Kenny, 1971; Bennett and Hacker, 2003; Bennett et al., 2007) The Wittgensteinian objector could observe, rightly, that one commits the mereological fallacy (the homunculus fallacy, alternatively) in attributing formal understanding and activity to any of the (material) parts or components (e.g., the brain) of the human being. For formal understanding is something that can only be intelligibly predicated of the *whole* human being, not of any of her *parts*—all of which are, of course, material. But it does not follow that therefore formal understanding is immaterial. For the human being is taken to be a *material whole*. And the whole, even when wholly material, may have capacities and abilities not attributable to, and not found in, any component parts. The mereological argument entails nothing about immateriality. Indeed, recognition that what is properly and intelligibly attributable to the whole need not (and the case of formal understanding and activity, *cannot*) be properly and intelligibly attributable to its parts does not lead to the claim that therefore what is attributable to the whole is somehow immaterial. What Kripke has shown, the Wittgensteinian objector could say, is that no component parts of a material whole can intelligibly be said to be the locus, agent, or source of formal understanding. But to conclude form this that the human being has

an immaterial faculty or capacity of some sort is not only to conflate mereological issues about whole and parts with metaphysical issues about materiality and immateriality, a categorical mistake nonetheless. Analogy: consciousness and voluntary movement is attributable only to the whole animal, not to any of its parts, but it does not thereby follow that consciousness and voluntary movement must be immaterial.

This is an interesting objection and bears some resemblance to a previously discussed objection which accused us of conflating issues of irreducibility with issues of immateriality. But it misses the point. In fact, it is simply irrelevant. We agree with the mereological fallacy, but it was not part of our argument that because no component part of the anything physical can determinately carry out a formal operation that therefore the operation could not be carried out at the system level. Our argument applies to anything that is wholly material, whether organic or inorganic, whether a whole or its parts, whether a mechanism or its subcomponents. What we say is that if the whole is a purely material whole, then formal understanding cannot be intelligibly attributed to it because it is not the sort of thing that could acquire the ability to think in formally determinate ways. So, we agree with the mereological fallacy: formal understanding is predicable only of the whole human being, the whole rational animal, but not on account of her material, neurobiological, endowment whether taken in whole or in parts. The capacity for formal understanding (or formal ability) cannot be wholly a material capacity, and neither can the ability (the understanding) and the activity consequent upon that understanding.

6. THE CONTENT FALLACY OBJECTION

We now turn to objections that can be raised in the context of Feser's appropriation of Ross's argument. We begin with an objection from Robert Pasnau (Pasnau, 1998) which Feser considers as a possible objection to his argument. It consists in accusing Ross of being guilty of what Pasnau calls 'the content fallacy'. The content fallacy is the fallacy of conflating facts about the content of thoughts with facts about the thoughts themselves, about "what shape or form out thoughts take in our mind." (Feser, 2013, p. 19) In other words, it consists in taking what is true of the content of a thought as true of the thought itself. But this is fallacious, says the would-be Pasnau objector, for one cannot infer from the fact that the content of a thought has such-and-such a property to the fact that the thought itself must have such-and-such a property. To say otherwise leads to absurdity. Before showing how the objection works, it should be immediately evident to the attentive reader that the content fallacy objection doesn't even arise for *our* argument (KRA) because the argument depends not on the *content* of understanding but on the *form* of understanding. What motivates KRA is not the fact that we can think about, say, addition, but that we can carry it out, that we can *add*. We made this clear from the outset (see chapter 1). Therefore, nothing in our argument relies on facts about the content of the understanding even though we granted that the structure or form of the thinking *could be* the content of some thinking (say, as when one thinks about conjunction). But one thing is *to conjoin*, another is to think *about conjunction*. Our argument against naturalism concerns, first and foremost, the former. Thinking (competently) about addition will, of course, not be naturalizable on our watch. But it won't be on account of the mathematical operation figuring determinately as the *content* of thinking, but rather on account of the fact that it is a kind or form of thinking and is therefore the pure structural outcome (e.g. adding) of the active exercise of understanding. For all we've said, content could be naturalizable. The reason why formal content (that is, competent thoughts and judgments about formal concepts, rules, and operations) is *not* naturalizable will not lie in the fact that a formal operation can figure as the content of a thought but rather in the fact that a formal operation *is* a form of thought or of the understanding that *can* be thought about by the sort of being that can understand such sorts of operations. Once one shows

that such pure forms cannot be materially realized in any determinate and intrinsic way, yet they can be thus realized in human formal understanding and activity, then we have an argument against the possibility of naturalizing formal understanding.[274] For all we've argued, there's nothing special about content as such.

The objection, however, does reasonably arise for Feser's *semantic* version of the argument. He has argued that the meaning of words or the content of thinking (i.e. concepts or conceptual content) is determinate in a way no physical thing can be and therefore cannot be wholly physical. From this he infers that the thinking itself cannot be wholly physical. His reply to Pasnau's objection is the following:

1) The objects of thought have property X, which entails that they are immaterial.
2) But thought itself also has property X.
3) So thought must also be immaterial.

Note that in order to avoid the content fallacy the property appealed to in premise 2) needs to be a formal one rather a content or semantic one. This is precisely the shift that Feser must recur to in order to meet the content fallacy objection: "…for Ross the X is determinacy … Ross can be read as saying that, just as the abstract form of inference *modus ponens* as studied in logic is determinate, so too do we have thoughts that are determinately of that form; and, just as the former could not be determinate if it were material, neither could the latter." (Feser, 2013, p. 20)[275] If by "we have thoughts that are determinately of that form" he means determinate in content, then the objection has not been answered. At any rate, the response is still fallacious if the conjunction of both premises is taken as the inferential grounds for the conclusion. For then, again, it would be arguing from (a property of) the content of the thought to (a property of) the thought itself. Otherwise premise 1) is superfluous and unnecessary, as the proponent of KRA would say, in order to establish something like 2) and 3). For, first you show that the physical cannot have property X. Then proceed to show that the cogitative has property X. Then conclude, *without any recourse to talk of content*, that the physical and the cogitative cannot be the same. I oversimplify, but this suffices to bring home the point.

Let us further consider the implications of Feser's response to the content fallacy objection by applying what logicians call 'analogical reasoning', that is, keep the *form* of the argument but change the content and see if a valid inference has actually been drawn. So, for example, take the following fictionalist variation on Feser's response:

F1) The objects of thought have property X, which entails that they are fictitious.
F2) But thought itself also has property X.

---

[274] Remember that we are not saying that pure formal structures are but forms of human understanding. That might be true or that might be false. What we are saying is, rather, that, whatever pure forms may or may not be, they are and have to be *at least* forms of human understanding. Otherwise human beings couldn't do the things that they do, like think logically and mathematically in virtue of the exercise of their cogitative powers. See the discussion in chapter 1.

[275] I must confess I do not understand, nor see the point, of the distinction made here between the determinate abstract form of inference known as *modus ponens* which is studied in logic on the one hand and us having thoughts that are determinately of that very same form on the other (that is, us being able to reason in accord with that form of inference). But a form of inference is, ipso facto, a form of reasoning. Plus, who does and studies logic? We do, of course—by 'we', I mean human beings. So this seems to amount to a case of a distinction without a difference. But perhaps Feser has in mind some sort of Augustinianism about forms of inferences, which are abstract and in the mind of God, but in which we participate in while reasoning (derivatively, of course). Note that KRA poses problems for this view if it makes our forms of inferences present in reasoning derivatively and extrinsically rather than in a originally and intrinsically. Anyways, this is somewhat of a bewildering distinction.

F3) So thought must also be fictitious.[276]

But this is absurd. We can try to save this type of argument by letting F3) say that the content of the thought is fictitious, as opposed to saying that the thought itself is fictitious (that is, a fiction). But then if having property X is sufficient to make whatever has X fictitious and, as F2) asserts, thought itself has property X then thought itself must be a fiction or fictitious. So, how is Feser's original argument more reliable than this one? In form, they are essentially the same. In the original argument, Feser intends premise 2) to be about the thought itself, not merely about its content or object. So if we take F1) and F2) as behaving, logically and argumentatively, in the same way as 1) and 2), and go on to deduce F3), then we have a textbook example of the content fallacy. True, Feser could response by saying that in the original argument there is only one variable and the predicate 'immaterial' in premise 1) is meant to remain constant and invariable. So the argument from F1) to F3) would violate this condition. Let's ignore issues of arbitrariness in this response and the fact that the argument, if it is a *type* of argument, *can* be schematized. Suppose we concede this point. Here is the problem: unless 1) is somehow relevant for the truth of 2), it is simply an unnecessary and irrelevant premise that, to add insult to injury, makes the argument vulnerable to the content fallacy since 1) like F1) is a straight forward appeal to the (properties of) the *content* of thought. If, however, premise 1) *is* relevant for the truth of either 2) or 3) (or both) then Feser is guilty of the content fallacy. In KRA, nothing like 1) should ever come up in an argument for the formal determinacy of understanding (cf. §7 for a further application of this line of reasoning). For what makes a cogitative act a formal (e.g. mathematical) act and therefore metaphysically asymmetrical in the relevant respect with the physical is *not* the content or object of the formal act (e.g., apples, the speed of light, sets, etc.) but the form (e.g., adding apples, calculating the speed of light, proving that a set is enumerable, etc.) of the very act, viz. the rule applied, the operation carried out.[277]

Our argument can be put this way,

(1) Intellectual activity can be formally structured and determinate.
(2) Whatever is formally structured and determinate cannot be wholly physical.
(3) Ergo, intellectual activity cannot be wholly physical.

But let's give Feser the benefit of the doubt. If by the statement "so too do we have thoughts that are determinately of that form" Feser means determinate in cognitional structure (in our jargon, determinate in formal structure), then Feser has effectively avoided the content fallacy. For no content-based reason is now given for the fact that thought has property X (presumably, the property of being formally determinate), a property the physical cannot have on account of physical properties alone. Unfortunately for Feser, his argument for the claim that thought is immaterial

[276] This is just one possible variant. Consider one that takes into account the fact that the objects of thought in question are mathematical and logical. So, interpreting the property X in premise 1) accordingly, we have: 'The objects of thought have the property of being mathematical/logical'). Suppose further that being a mathematical or logical object entails that it is something eternal or sempiternal. Then, according to this scheme of argument, thought itself would be eternal or sempiternal. Absurd. The scheme puts no constraints on what property X could be.

[277] I am not saying that no argument from (conceptual) content or meaning against naturalism can be given. I am saying that the argument from formal determination against naturalism is not an argument from content. That's it. True, given KRA, formal content won't be naturalizable. But formal content is nothing but formal rules, functions, and operations figuring as objects of a judgment or thought and which we have, independently of their status as possible contents, shown not to be naturalizable. Whether (conceptual) content as a whole is naturalizable an open question. For an argument against naturalism properly characterized as an argument from content (truth), see e.g. (Plantinga, 2011)

relies heavily on the fact that formal thinking is *semantically determinate*, which is a content-based grounding for an inference to the immateriality of thought. But the special feature about formal thinking is not, for Ross, it's content but it's *form*.[278]

You will have noticed that, unlike Feser, we do not talk about 'thoughts' or 'conceptual content' in developing KRA but rather about 'understanding' and therefore about the tripartite nature of the cogitative: the capacity, the ability (or abilities), and the act or activity. Pure functions can be seen as rules for determinate sorts of activity with certain formal ends, be it calculating, measuring, inferring, proving, etc. We have not argued from the content of the activity or the understanding to the metaphysical nature of the activity and understanding as such. We have, rather, argued from the form (structure) of the activity and understanding to the metaphysical nature of the activity and understanding as such. We have followed the *formal* route rather than the *semantic* route (Feser, 2017, 2013).[279]In short: the reason why the content fallacy cannot even arise for KRA is because the argument from formal indeterminacy is not an argument from the content of formal thinking but from the *form* of formal thinking. Of course, once one establishes that the forms of understanding cannot be a wholly physical and therefore that the invariable components of such a kind of understanding and its consequent activity are not physically determined one can then make the case that whenever one thinks *about* a given pure form or possess a formal concept, the thought or the concept is about something non-physical that we have independently shown to be, at least, a form or a pure structural principle of understanding. Hence, even Ross's parallel argument as to the content of understanding is not subject to the content fallacy so long as it is seen as a corollary to the argument from the form of understanding. Thus, 'formal content' refers to our concepts or thoughts about pure forms, which are the structural ways of definite formal intellectual activity.

Note also that issues of semantic determination are conceptually interconnected with issues of semantic individuation (see e.g. Burge, 2007, cf. 1979). I suppose that Feser would be, partly at least, an anti-individualist and anti-internalist about the individuation and determination of the semantic content of our true thoughts about the world insofar as we can acquire the concepts of and make true judgments about material, natural substances like dogs, water, gold, etc. Feser may also want to give an anti-individualist account of the concepts of formal objects like numbers, say numbers as mere abstractions from the multiplicity of discrete material units we encounter in the world and thus determined in part by features of the world we inhabit, letting such things a complex numbers, quaternions, Riemannian geometry, Aristotelian diagrams (see e.g Demey and Smessaert, 2016), etc., be but ever higher and further abstractions from, ultimately, a set of upmost basic abstractions from

---

[278] There is something to be said about content, though, but on Ross's own explanatory framework what has to be said goes in an entirely different direction than Feser's. An example of this is that, according to Ross, the non-formal content of any judgment (say, judgments about *reality*, about matter, about the universe, about social institutions, about particular happenings and events, etc.) will lack an intrinsic feature of the physical, namely what Ross calls 'the transcendent determinacy of the physical'. The array of material and natural conditions, and necessities, that must obtain to make a given judgment about a particular reality true (it's truth-conditions and truth-makers) overflow indefinitely anything that that the judgment could possibly contain. So Ross (see e.g. Ross, 2008) gives an argument for the inverse of KRA: there are properties that the physical has that no judgment or thought could possibly have. There is something unique and special about the physical also, and human judgmental understanding cannot even come close in capturing it.

[279] Indeed, even if we concede Chomsky the premise that one need not grasp, know, or think about a formal rule or pure function in order to follow it or compute it, our argument would still hold whereas Feser's wouldn't. As I've already emphasized, since our argument concerns a unique set of abilities and the exercise of that set of abilities (e.g. to compute a mathematical function) and can be made compatible with the claim that one can compute pure functions without knowing that one is doing so, there would still be formally determinate activity even if no semantically determinate thinking in the sense Feser understand this takes place. I do not, of course, think there can be 'rule-following' without 'rule-understanding' where the rule is a formal rule, but that is a claim to be defended elsewhere (see, however, chapter 1).

the natural order.[280] But there are two problems. First, what determines and individuates semantic content? Second, how is it not the case that *any* material structure will turn out to be indeterminate with respect to our concepts of it if it is the case that concepts or conceptual contents are semantically determinate and in virtue of *that* immaterial? Feser's semantic version of the argument seems to be unable to answer in a non-*ad hoc* manner why it is the case that when the semantic content is of a pure function matter cannot be determinate with respect to the concept but when the semantic content is of a material function matter can be determinate with respect to the concept (as we will see, this is precisely the basis for Dillard's 'schmitosis' objection to Feser's Ross, see §8).[281] Take the concept of photosynthesis. Can't there be a qu-like biological concept, say qu-photosynthesis, that is incompossible with the biological concept of photosynthesis yet is satisfied equally well by the biological process under investigation? After all, if the semantic considerations laid out by Feser can apply to color concepts and expressions (as Goodman, 1983 showed and Feser himself acknowledges), why can they not apply to biological concepts and expressions? How can Feser rule one or the other out? We will come back to this later when we consider Dillard's objection.[282] For now, just note that it is not clear how Feser can escape the content fallacy except by shifting from the content of the thinking to the form of the thinking. But this is effectively to abandon the semantic version of the argument. Ross's argument, or KRA at least, concerns primarily the latter and is therefore not subject to the content fallacy objection.

7. PUTNAM'S OBJECTION

---

[280] It is not clear what is the content of the lowest and original set of abstractions (which are not abstractions from other abstractions but abstractions from the physical features of the substances and processes of the world) that constitutes mathematics but maybe it consists of Euclidean geometry, the rational and irrational numbers and the standard arithmetic operations of addition, subtraction, multiplication and division.

[281] Formal and non-formal thinking can both be semantically determinate, and formal thinking can be just as much about non-formal objects as non-formal thinking can be about formal objects (e.g. compare adding apples or deducing the existence of dark matter from some observations and experimentation to pondering Riemann's hypothesis or thinking about *modus ponens*). Of course, maybe all thinking is formal under some broader sense of 'form' (see Ross, 2008, p. 123 for the claim that what applies to mathematical and logical thinking applies too all thinking though it is harder to make convincing for non-formal thinking). "That all thought is determinate in that way [determinate amongst incompossible pure forms] is harder to make convincing because it rests on one's recognizing that whatever thinking we do, whether simple assertion, or hoping or wanting or intending (over the whole family of things each of those can be, according to its particular content on a particular occasion) is in *form* such that, in order to do that, we have to do what would be the same activity for indefinitely many other cases (sorted by content) that do not happen. Further, someone else might have thought, or said, or believed, or felt the same, in a way definite among incompossibles, though not in a way numerically identical with your or my thought, or even with a directly similar brain state or feeling. Thinking always has "other cases" as to form—I can think the same thing again, or remember it, for instance—though the form may not be separately articulable by us the way mathematical and logical forms are."

[282] In the case of material substances and formal objects, Feser would presumably say 'nature, that is, the material reality' individuates and determines the semantic content of material concepts (this has to do with Feser's commitment to the Aristotelian notion of 'formal identity', see n. 283). But what about formal operations? There is no more addition in nature than there is addition when leaves pile up. Our argument should have made it clear by now that even if formal objects can be said to be abstractions from the natural order in a literal way, formal operations cannot be said to be such abstractions from material process and functions in a literal way (that is, as if these operations were somehow there in matter and we abstracted from it what was already there). I am of the opinion that 'abstraction' has to be analogous rather than univocal when used in respect to material structures and when used in respect to pure formal structures. (Numbers can be defined set-theoretically. If this is what constitutes numbers, then numbers cannot be said to be literally abstracted from material reality. But this is an open question and is an issue best reserved for the philosophy of mathematics.)

What I call 'Putnam's objection' is in fact a host of several interesting objections raised by the late philosopher Hilary Putnam (Putnam, 2002) in response to John Haldane's Aristotelian doctrine of 'formal identity'. Though not raised as an objection to KRA, it is interesting to note that KRA places constraints on the kinds of plausible answers that might be given to Putnam's objection. Plus, it is an objection that Feser himself considers and attempts to meet elsewhere, though (as we shall argue) at the cost of his semantic argument for the immateriality of the human intellect. So, we consider it here. Recall Feser's response to the content fallacy objection,

1) The objects of thought have property X, which entails that they are immaterial.
2) But thought itself also has property X.
3) So thought must also be immaterial.

Suppose we consider a formal *object* like a triangle and let that be the object of thought under consideration. Now let us rephrase the argument as follows,

1´) The object of thought has the property of being a triangle, which entails that it is immaterial.
2´) But thought itself also has the property of being a triangle.
3´) So thought must also be immaterial.

But Putnam finds this absurd, and rightly so.[283] The thought of a triangle is not a triangle, and it is hard to see what sense one can make of such an argument. So, what is the difference between a thought of a triangle and a triangle? Notice that Feser is quite selective, for it is not clear why a predicate variable cannot take the place of 'immaterial', to wit:

---

[283] To be sure, Putnam's objection has to do with 'dogness' considered as the form (in the Aristotelian sense of the term) of a dog and as the form (read: content) of the thought, concept, or true judgments of a dog. The form 'dogness' must be had by both. This what Aristotelians call 'formal identity'. The form 'dogness' exists in the natural, material order of being (*esse naturale*) but it can also exist in our thoughts or true judgments about the essence of dogs (*esse intentionale*), and this formal identity is what is supposed to explain how our thoughts can be about dogs themselves. (Needless to say, this Aristotelian notion of 'formal' is not univocal with our use of the notion of 'form' and 'formal'.) It is supposed to explain how we can grasp the essence of things and how our mind really grasps what is 'our there' in the world. In the case of particular dogs, 'dogness' is materialized, whereas in thoughts, it is not. That's why the latter do not become dogs yet are nevertheless constituted by dogness. Here's the passage from Haldane that Putnam is reacting to:

> [W]hat I find missing from Putnam's discussion . . . is any explicit account of how this is possible. [The 'this' refers to the statement, quoted immediately before, that 'successful perception is just a seeing, or hearing, or feeling, etc., of things 'out there' and not a mere affectation of a person's subjectivity by those things'.] Elsewhere I have urged the merit of the maxim 'no epistemology without ontology' . . . if there is to be the sort of conformity of mind to thing which Putnam and McDowell seek, then I can only see this being provided according to an account of the sort developed by Aquinas when he writes that the intellect in act is the intelligible in act or less scholastically, that mind will only be of a thing when it is formally identical with it; when what we think and what is thought are the same … What does this mean? And how is it possible? It means that when I think of something, that which makes my thought to be the kind of thought that it is – a dog thought, say – is formally identical to that which makes the object of my thought to be the kind of thing it is, a dog . . . The form of dog exists naturally and substantially (*in esse naturale*) in the dog, and intentionally and predicatively substantially (*in esse intentionale*) in the thought. (Haldane, this volume, 101-2)
>
> (Putnam, 2002, pp. 105–106)

It is Feser who introduces the 'triangle' example in his response to Putnam. But this backfires and points to a huge blindspot in Feser's overall metaphysical approach to formal objects and human thought, in particular with how his reply squares with his semantic argument for the indeterminacy of the physical.

1´´) The objects of thought have property X, which entails that they have property Y.
2´´) But thought itself also has property X.
3´´) So thought must also have property Y.

Suppose we let X be 'being a triangle' and Y be 'being triangular and trilateral' or 'being consequent on Euclid's axioms'. Then the argument would entail not only that the thought of a triangle is itself a triangle, but logically that the thought itself is triangular and trilateral and consequent on Euclid's axioms. A host of other counterexamples can be found to illustrate the absurdity of this approach.[284]

Now, Feser has a reply to Putnam's objection (Feser, 2016), but it contradicts his argument for the formal indeterminacy of the physical.[285] Recall that Feser's version of the indeterminacy argument applies just as much to formal objects (e.g. geometrical figures or numbers) as to formal operations (e.g. addition or commutation), though taken as content. Feser's reply is to say that the second premise is false because something can only be a triangle when it is *materialized*. But the form of triangle in thought is not materialized. Hence, the thought of a triangle is not triangular—in the sense that the thought of a triangle does not become itself a triangle. The difference between the thought of a triangle and a triangle is that the latter is a material object and thus a materialized entity, whereas the former isn't and therefore cannot become a triangle, since for something to be a triangle it must be a material figure of a determinate triangular shape. Feser writes,

> The form being a triangle – or, as I put it above, being a closed plane figure with three straight sides – does not suffice to make something a triangle, because both triangles and thoughts about triangles have that form, and the latter are not triangles. Being a triangle requires both the form in question and matter.
>
> (Feser, 2016)

Feser makes other statements similar to this one throughout his post. There problem is that the criterion he gives for what makes something a triangle is inconsistent with the claim that material objects are formally indeterminate (a claim which Feser defends and is committed to). If Feser's Rossian argument is sound, it entails that geometrical forms (as well as, for example, arithmetical, logical and algebraic forms) cannot be determinately realized by any material object; thus, being intrinsically and genuinely of such a form cannot be a matter of having matter—since, again, there is no such thing as matter being a determinate instance of, i.e. intrinsically and determinately realizing, an abstract form. Matter can only approximate such forms, but it cannot be of such forms, which is why no material object can exclude its being a case of incompossible abstract forms. With sufficient ingenuity and imagination, one can construct an incompossible geometric form, say 'qutriangle', which the material 'triangle' equally well satisfies. So, what it is to be a triangle cannot be a matter of conjoining pure form and matter.[286]

---

[284] To be sure, in his response to Putnam, Feser does not use the argument schemata he used in response to the content fallacy objection. I am using it to showcase the sort of reasoning behind Putnam's objection, and the resemblance between Pasnau's content fallacy objection and Putnam's objection.

[285] The reply was posted in a blog post (http://edwardfeser.blogspot.com/2016/05/putnam-and-analytical-thomism-part-i.html). I replied immediately to in the comment section of that blog post (see the comments by A. R. Díaz). Here I have merely restated the objections I raised there.

[286] But you might say, "Perhaps Feser has something like 'intelligible matter' in mind. This is an Aristotelian notion to which your claim about indeterminacy does not necessarily apply." Perhaps, but then it would not be a response to Putnam's objection. It would be an *ignoratio elenchi* on the part of Feser. See our argument on the next page.

In other words, there is a categorial gap between the abstract form and the matter that is supposed to realize that abstract form. Ross's argument, which Feser himself has defended, entails that no kind or amount of matter can claim as its own a unique and definite abstract form. To sum up, what it is to be a triangle cannot be a matter of materially realizing the abstract form 'triangle' since no such thing is possible—except, as Kripke says, in an observer-dependent (i.e. extrinsic) way. But if this is the case, then the criterion he gives, on the basis of matter, for differentiating what it is to be a triangle from what it is to be a thought of a triangle collapses, since materialization is not sufficient to constitute the former—indeed, it cannot be. Feser knows this well. Indeed, in his response to objections raised by Peter Dillard (Feser, 2013, pp. 22–23), he replies by saying that to suggest that mathematical and logical forms can be determinately realized by material things but that we don't know it is as silly as saying that geometrical forms are determinately realized in material things but that we just don't know it. So he knows (or, rather, is strongly committed to the claim) that material things cannot be of a definite abstract form, whether this be geometrical, logical, algebraic, etc. Perhaps they can be extrinsically determinate (say, by reference to an observer, as Kripke argued), but this just means that in themselves, i.e. intrinsically, they realize no such forms; only relative to us, to our interpretative conventions and practices, our interests, or what have you, can they be said to realize any abstract function at all. But then this would be true of his criterion for what makes a thing a triangle (rather than a thought about a triangle).

Feser can respond by saying that triangular things do exist (e.g., traffic signs, pyramids, etc.) but they exist only as approximation to the geometrical form of triangularity; therefore, his reply to Putnam must be understood as stating that the difference between the thought of a triangle and a triangle is that the latter is a material (and therefore imperfect and inexact) exemplification of the concept of a triangle. If this is his response, then all well and good[287] (perhaps this is what 'intelligible matter' means for the Aristotelians when talking about the materialization of pure formal structures). But then Putnam's objection remains unanswered. Putnam was not after the metaphysical differentiation of the thought of a triangle from the approximation or imperfect realization of a triangle (which is what any materialized or material triangle is) but rather of the thought of a triangle from a triangle as such. Feser's criterion for what it is for something to be a triangle was not in terms of the "everyday, practical, triangularish sense" (which is nothing but a placeholder for 'triangle in an approximate and observer-dependent way') but rather in terms of what makes an object, intrinsically and *in se*, a triangle (as opposed to a thought of a triangle). Otherwise, if he were using it in the "everyday, practical, triangularish sense", that is, in the sense that material objects though recognized to be imperfect realizations of pure formal structures are nevertheless for practical purposes said to be of definite geometrical forms, then he would not be answering Putnam's objection, which was that Aristotelian-Thomistic (A-T, for short) metaphysics does not have the resources to distinguish between something's being, intrinsically and really, a triangle from something's being a thought of a triangle. Feser answered that it did have the resources because what it is for something, intrinsically and objectively (i.e. in an observer-independent way),

[287] Except for the fact that Putnam's objection originally concerned the difference between 'being a dog' or 'dogness' (what it is for something to be a dog) and 'the thought of a dog' (what it is for something to be a *thought* of a dog rather than a dog) as opposed to the difference between 'being a triangle' and 'the thought about triangle'. So, this maneuver would only constitute a response to Putnam's objection if it could be applied to 'being a dog' or 'dogness' and 'the thought of a dog'. But applied to dogs this maneuver turns out to be manifestly implausible, especially for an A-T philosopher like Feser, since it would imply that dogs are approximations, imperfect realizations of 'being a dog' or 'dogness'. Dogs would be indeterminate with respect to what it is to be a dog, so there wouldn't be dogs at all (at least materially) except in an observer-dependent way. This is surely absurd. So, a point in our favor: what can be said of geometrical structures, and pure formal structures in general, cannot be said of dogs, and material structures in general, and conversely.

to be a triangle is for the geometrical form to be realized in matter. But this is inconsistent with his version of Ross's argument. Therefore, given the formal indeterminacy argument there is no such thing as a material triangle, intrinsically speaking, only approximations and imperfect realizations of the given Euclidian geometrical figure. Triangularity is a mind- or observer-dependent feature of material things (granted suitable material conditions and organization).[288] So Feser faces a dilemma: either he abandons the semantic-formal indeterminacy of the physical so as to allow that what distinguishes a thought of a triangle from a triangle is that the latter is physical or physically realized, or he abandons his response to Putnam, in which case Putnam's objection remains unanswered.

Now, how is all of this relevant to KRA? The answer is simple: because depending on how one interprets KRA (particularly, Ross's argument for KRA's second premise), one will be constrained in the set of plausible explanations that can be given to the nature of the materialization of pure formal structures. KRA implies that pure formal structures an only be imperfectly realized by anything physical, and therefore material reality only approximates pure formal structures. So, what it is for something to be a given pure formal structure cannot be a matter of matter. In other words, what it is to be a pure form cannot be a matter of materialization. But what it is for something to be an approximation, an imperfect realization, of a pure form *can* be a matter of matter or materialization. Since Feser accepts a semantic version of KRA and has repeatedly acknowledged that geometrical forms cannot be realized in matter determinately, his answer to Putnam's objection is not to be point since it does not give an account of what it is to be a triangle (as opposed to an imperfect realization of a triangle). This is why I belabor the point that in his response to Dillard, Feser claims that what applies to geometrical objects or figures applies to mathematical and logical forms (say, functions) in general and that what applies to geometrical figures is that (1) they cannot be determinately realized in matter and (2) we know this. This is why Feser responds by saying that to claim that mathematical functions might be realized determinately in matter but that we just don't know it (which is the claim being made by Dillard) is as silly as suggesting that geometrical objects or figures, say, a triangle, is determinately realized in matter but that we just don't know it. For we *do* know that matter doesn't realize any such geometrical figure or form determinately, but only imperfectly and approximately. (On Feser's own account, the argument concerns both mathematical operations and mathematical objects. Perhaps Feser would say that it does not concern all mathematical objects, but only a subset of them, namely, geometrical ones. But that is sufficient for my objection to Feser. In addition, he would also have to give an account as to why it is applicable to only these and no other mathematical or formal objects.)

In my opinion, Feser is right in saying that the formal indeterminacy of the physical concerns all mathematical and logical phenomena, and that Ross's argument can be shown to be extendable to all of them—even to music scores, albeit with qualifications. (And mathematical and logical objects are *per se* indeed a problem for materialism, since one can argue that if we were entirely material, we would not even be able to think and grasp them, let alone carry out determinately functions over them. Also note that by 'abstract' I do not mean 'Platonic'. I am leaving this question open, although I am inclined to think that both the Platonic and Augustinian ways of understanding the metaphysical status of mathematical 'objects' are mistaken. There's also the further question of what exactly the relationship between mathematical objects and mathematical functions/operations is, but that's a huge and difficult issue.) Yet, as I've repeatedly stressed, Ross is arguing primarily about

---

[288] Again, if understood in terms of our Kripkean argument, this is not a triviality argument. Not everything (that is, every material thing) counts as a triangle. But whatever counts as a triangle counts equally well as a 'qu-triangle'. Judgments about the geometrical structure of material objects is grounded on the material structure of the material objects in question but not determined by it. Again, a particular material organization and structure will be necessary for formal determination in matter but not sufficient. The sufficiency conditions that make geometrical judgments true of particular material objects are extrinsic to the material objects themselves.

mathematical and logical *operations* (e.g., reasoning in accord with the rule of constructive dilemma) and so are we. Thus, we leave it open whether mathematical entities or objects as such are nothing more than products of human understanding and productive intellectual activity. We do not, however, leave it open whether they can be present in matter in a mind-independent way. Hence, if what KRA says about formal operations is *mutatis mutandis* true of mathematical objects (as Feser thinks it is), namely that matter cannot determinately realize and be an instance of any mathematical object on account of its material properties alone, then one cannot answer Putnam's objection about the possible collapse of thoughts about triangles into triangles by appealing to material conditions of any sort as these will satisfy at the relevant level of abstraction figures and descriptions of incompossible geometries. So, the relevance of Putnam's objection to KRA is twofold: one, KRA places constraints on the set of plausible answers to any Putnam-like objection and, secondly, it strengthens our critique of Feser's appropriation and incorporation of Ross's argument into a conceptual framework in which Ross's argument does not fit as neatly as Feser presumes.

Let us now state the obvious: nothing in our argument makes KRA subject to anything like Putnam's objection. This is because KRA and A-T need not be bound together. Putnam's objection and Feser's reply exposes a tension between Ross's argument for the formal indeterminacy of matter and Feser's A-T approach to the nature of geometrical objects which (at least if his response to Putnam is to be successful) requires that matter be formally determinate (that is, that matter be determinate with respect to forms which Feser has in his Rossian argument denied can be determinately realized in matter!). In one sense, it is unsurprising that Feser rephrased Putnam's objection in terms of a geometrical figure, as it makes the purported absurdity of the A-T approach more readily apparent. But in another sense, this is quite surprising given that he has previously argued for the claim that geometrical figures cannot be determinately realized in matter. Plus, Putnam's objection can be raised using the addition function instead. He could've asked, 'on the A-T approach, what distinguishes a thought of the addition function from the addition function itself?' Surely, Feser would not (or should not) respond by saying that the difference is that the addition function is in matter or is materialized (say, that it is a material function!) whereas the addition function as the content of thought isn't. He can't say this if he hopes to uphold his Rossian argument against the formal determinacy of matter. There is no addition or addition function in matter, except denominatively and derivatively. But it seems like his peculiar brand of A-T would *force* him to say this; if not, he owes us an explanation as to why his response is acceptable in the case of geometry but not in the case of arithmetic. We, on the other hand, face no such problems. For our argument concerns only formal operations, the application of formal rules and the structure of intellectual activity. We do not make and have not made the case with respect to formal objects (if such a thing exists) and our thoughts about formal objects, except perhaps the generalizable point that the formal (whether entitative or operative) cannot be present intrinsically and determinately in the physical. Furthermore, it is indeed a tricky road to infer something about thoughts themselves on the basis of the possible contents of thoughts. This is alien to our approach. We've concerned ourselves with the ability to follow formal rules, whether in computing a certain mathematical function or in thinking in ways that accord with deductively valid forms of reasoning. In the case of formal operations, therefore, it is true that the formal structure in question constitutes the kind of intellectual act or activity. So, to Putnam's objection, we have the following response: if it concerns formal objects, KRA is silent;[289] but if it concerns formal operations, KRA says that the difference

[289] Though not entirely if it turns out that what is true of formal operations is true of formal objects in its entirety. KRA can equally show that whatever one considers a formal object (e.g., a geometrical figure or a number) to be, such a thing cannot be determinately realized in matter. Yet KRA doesn't have much to say about what makes a formal object a formal object. Indeed, maybe KRA and Platonism are ultimately compatible, despite appearances to the contrary. So,

between the addition function itself and a thought about addition is that the former is *at least* a form (a structural principle, a normative invariable component, a rule) of thinking or intellectual activity whereas the latter is a form of thinking figuring *as the content* (the variable component) of some thinking or intellectual activity. That's it. And since we've shown that the pure forms of understanding cannot be intrinsically and determinately present in matter, yet they are thus in the understanding, the presence of such pure forms in the understanding cannot be wholly a matter of matter. In sum, our discussion of Putnam's objection should make it clear that KRA need not be committed to the Aristotelian notion of 'formal identity' as understood by Feser in his response to Putnam. There is no tension.

8. DILLARD'S OBJECTION: MITOSIS OR SCHMITOSIS?

This objection was briefly addressed in chapter 3. In his criticism of Ross's argument, Dillard objects that Ross's argument only establishes that it is epistemologically underdetermined which function a physical process realizes and that we have been given no grounds for concluding that it is metaphysically indeterminate which pure function a physical process realizes. We dealt with the epistemic objection above (see §1). I want to focus now on a counterexample to Ross, proposed by Dillard, that if correct would seem to undermine Ross's argument.

> Consider mitosis, in which the nucleus of a cell divides through the phases of prophase, metaphase, anaphase, and telophase to yield two new nuclei, each containing an exact copy of the parental nucleus's chromosomes. Mitosis is a purely physical process, and the replicated chromosomes are a paradigmatic example of what Ross calls "definite natural structures" as opposed to "pure functions." However, the distinction Ross draws between natural structures and pure functions is far from obvious. Just as in the case of a calculator and the addition function, the cells of a growing organism equally conform to incompatible natural structures. Suppose that these cells have undergone 10^n mitotic divisions. Then the cells' behavior equally conforms to mitosis and to "schmitosis," where the latter process yields nuclei containing an exact copy of the parental nucleus's chromosomes for the first 10^n cell divisions but an entirely different set of chromosomes for any cell divisions > 10^n. No matter how many mitotic divisions the cells undergo, their behavior will also conform to an incompatible, non-mitotic process. Must we infer that it is metaphysically indeterminate whether the cells are undergoing mitosis as opposed to schmitosis?
>
> (Dillard, 2014, p. 140)

The answer is obviously 'no!'. Why? Here Dillard and I part ways. He thinks that the reason why we must not infer such a conclusion is because what follows "is simply … that more complex physical facts about mitosis are not identical with facts about cellular behavior or more basic physical facts about energy, gravity, subatomic particles, atoms, and inorganic molecular compounds. In other words, microbiology is not reducible to physics and inorganic chemistry." (Dillard, 2014, p. 140) He

---

how would we cash out the nature of formal objects like geometrical figures and diagrams, numbers, groups, and so forth? Surely, these cannot be considered 'forms of understanding' in the same sense in which functions, rules, or operations can be thus considered? Formal rules can be applied to formal objects (e.g., adding real numbers), so surely there is an asymmetry between formal operations and formal objects? Or maybe the latter is reducible to the former? How could this possibly work with respect to, say, the fact that there are infinitely many transfinite cardinal numbers that keep getting bigger and bigger without end (Balaguer, 2018)? There seems to be plenty of formal truths that seem to be independent of human formal judgments and operations. We leave this open, as this is a matter for metaphysics and the philosophy of mathematics.

goes on to say that the indeterminacy is only epistemic, and therefore is a case of underdetermination rather than metaphysical indeterminacy. Relative to our finite past observations, the biological process satisfies both mitosis and schmitosis, but it doesn't follow from this that no biological process at all is taken place nor that the cells aren't undergoing mitosis or that if they are, mitosis must be a non-physical process.

But this response seems to me to be an *ignoratio elenchi*. First, we are in agreement that microbiology is not reducible to physics and inorganic chemistry. But microbiology will be just as formally indeterminate as physics and inorganic chemistry. Formal indeterminacy affects matter at every level. Second, if the process is a biological process, then it is of course *not* biologically indeterminate. A biological organism or process is precisely a biologically determinate substance or process. Otherwise there would be no biological organism or process at all! There can only be conformity of the biological to incompossibles structures at the same time and in the same respect if the structures under consideration are of a different ontological category than that of biology. What Dillard fails to see is that 'schmitosis' is if anything an idealization, an abstraction, and therefore an abstract functional description of a given process. It is therefore subject to Kripke's argument. One can, of course, give formal, abstract functional specifications of mitosis too (all sciences use formal tools to model and classify natural structures) but then mitosis as a biological process cannot be identified with mitosis as the formalization of that process. Dillard is actually conceding Ross's point: these are idealizations, and relative to idealizations biological process will satisfy incompossible ones. That is why biology is not a pure formal science and not merely a matter of mathematized idealizations and descriptions, otherwise there would be no empirical component to the knowledge of biological reality. But matter at whichever level will satisfy incompossible pure functions and forms. Dillard wants to have his cake and eat it, but we need not concede as intelligible the claim "the cells of a growing organism equally conform to incompatible *natural* structures". What Dillard should've said is that "the cells of a growing organism equally conform to incompatible *abstract* or *formal* structures" and therefore being a biological organism or process cannot be not a matter of conforming to, or being an instance of, a definite abstract or formal functional structure (in the case of 'mitosis/schmitosis', a matter of satisfying an abstract functional description). There are physical facts about mitosis, yes. But these are precisely biological, not formal (e.g., functional abstract idealizations).

Dillard seems also to want to suggest that there might be higher-level, perhaps *sui generis* physical facts that are irreducible to any known level of matter which somehow determine pure formal structures: "… it does not follow that there are no physical facts about mitosis at all, but only that in addition to physical facts about our observations, subatomic particles, atoms, and inorganic molecular compounds, there are also non-reducible facts about mitosis … [and so can the materialist reply that] the physical fact that a human being or a calculator is adding rather than quadding is not identical with more basic facts of physics, chemistry, biology, and neurophysiology." (Dillard, 2014, pp. 140–141) The first sentence is incontrovertible, but the second one is astonishing. The first one amounts to the claim that biological processes are not reducible to non-biological processes and facts. Fine. We actually concede this. We are not reductionists. But the second claim is disanalogous, for it says that higher-level physical facts are not reducible to lower-level physical facts and that the former can be identified as the realizers of pure functions. But this just states the obvious about irreducibility only to fallaciously assume that there can be a kind of physical fact, i.e. a higher-level one, that can be identified as a definite realizer of a pure formal structure. There is no argument here, just an appeal to a conceivability thesis of some sort. What our Kripkean argument shows, however, is that such a thing cannot obtain. It is not about higher-level vs. lower-level physical facts, it is about physical facts *tout court*. It concerns the category of material substances and processes. And since the content of 'physical' and 'material' is supplied but the actual and potential levels of matter

(e.g., subatomic particles, atomic and molecular substances, inorganic chemistry, biology, neurophysiology, observable animal behavior, geology, etc.) Kripke's argument shows that none of them can be of a definite pure formal structure. But since they are definite in their respective structures (e.g., molecular if a molecular compound) yet indefinite with respect to their formalizations, being a natural structure is not a matter of implementing a given formal, abstract functional description or structure.

Two further problems. First, Dillard's reply seems to commit the materialist to a *sui generis* physical fact response. If so, it follows that our point is proven in that something *sui generis* is needed to determine a definite pure formal structure or pure function. But since it is also not clear what sense 'physical' has in '*sui generis* physical fact' then in the absence of a clarification, our argument in chapter 3 against the *sui generis* reply holds. But also, what are these higher-physical facts that transcend all other kinds of physical facts? How are they different in *kind* from any other, "lower-level", physical facts such as to determine a definite mathematical or logical structure? The materialist reply gives no example of, no content to, and no proof about such purported "higher-level physical facts" to which even neurobiology and neurophysiology are considered "lower". Likewise, we are never told how these physical facts could determine a pure function. If it is physical, all one needs to do is run Kripke's argument and indeterminacy will apply. Now, Dillard might just respond by saying that what would apply so such 'higher-level' physical facts is 'underdetermination' not 'indeterminacy' and that this has been his point all along. So, the 'higher-level' physical facts are underdetermined with respect to pure functions just as much as any 'lower-level' physical fact is. But why is this so? Also, suppose we found a candidate higher-physical fact but that it turned out that it is likewise formally underdetermined. Couldn't Dillard's materialist reply that there must or could be an even higher-physical fact still that can determine a pure function? How is Dillard's materialist not on the road to an infinite regress and an unfalsifiable proposal? What would it count for a proof or argument to be a metaphysical prove or argument? Dillard never tells us, so it is hard to know what would falsify Dillard's materialist.

Be that as it may, we have a simpler and more straightforward response. Dillard's position is manifest in the following passage,

> Our total physical theory of the world consisting of physics, chemistry, biology, neurophysiology, and additional physical facts about thinking may be under-determined by all possible empirical evidence, since an alternative total theory (perhaps one according to which we are quadding instead of adding) might be equally compatible with the same body of evidence. Yet no metaphysical conclusions about the indeterminacy of any physical facts follow from this purely epistemological basis.
>
> (Dillard, 2014, p. 141)

Let indeterminate be the metaphysical notion and underdetermined the epistemological one. The point is that *our* formal activity is *not* underdetermined (if it were, there would be no way of knowing and understanding what addition and quaddition are, and therefore no argument to be made, not even an underdetermination one! Quine (Quine, 1976) could not have even made his case for the global under-determination of our total scientific theory by empirical evidence, since there would be no telling which 'scientific theory' we are talking about). So, our formal understanding is *not* underdetermined. What does that say about *it* if any kind of physical fact, process, or object whatsoever will be formally underdetermined? For a fuller treatment of the epistemological objection, see §1.

To be fair, Dillard's confusion is not his entirely his fault. It springs from Feser's own distorted presentation of Ross's argument as one merely about the content of concepts or meanings.

What Feser's version of the argument obscures is that addition and quaddition are *operations* or *acts*, formal rules for *doing* something (say, counting). There is no such parallel with 'green' and 'grue' (Goodman, 1983; cf. Hacking, 1993) or 'mitosis' and 'schmitosis' except when we come to the *use* of such words or to their concepts. But pure functions are forms of understanding, structural principles that constitute certain kinds of intellectual acts and activity. Not all is lost, however, for an argument from meaning and semantics. Consider the following symbolic representation of a machine function for an endless loop to print the cube of every number starting with the number zero (Ross, 2008, p. 198 n. 14):

10 Z = X*X*X
20 Print Z
30 X = X + 1
40 GO TO 10

So, suppose we begin with the number zero such that at 10 'X = 0'. Then the value of Z at 20 will be '0' since the cube of zero is zero. At 30 we get the machine to add 1 to the current value of X, hence 'X = 0 + 1 = 1'. At 40 the machine goes back to 10 with the new value of 'X' and begins the process again of printing the cube of the new number. One can see that it is a machine function for endlessly computing and printing successive cubes. But these marks on the page above must be interpreted to constitute symbols of the relevant functions and for the to be understood. By themselves and absent an explanation of how they are used, they mean nothing. Nothing in the sequence of such letters and numerals entails determinately one function rather than another. Even calling them 'letters' and 'numeral' already requires an interpretation, linguistic and symbolic conventions, and practices where agreement on linguistic and formal notation exists. So, there is semantic indeterminacy. But there is formal indeterminacy too. Suppose we settle what those letters mean. Then we can say that a physical machine is computing successive cubes. But nothing about what the machine does excludes in and of itself an alternative incompossible interpretation and therefore its being a case of some other incompossible computation. In fact, even if we leave the notation uninterpreted, we can appreciate a *causal* physical pattern in the manipulation and transformation of physical vehicles (in the present example, physical vehicles that constitute symbols if interpreted, as in the case of 'X' written above). Yet such a physical causal pattern will be compatible with incompossible *formal* rules for manipulating and transforming such physical vehicles. Thus, there is indeterminacy not merely as to what the physical vehicles (e.g., the letters and numerals) represent but as to what it's being done over them or with them. "You can see that no matter what outputs the machine gives, it might have been doing something till then other than printing successive cubes, unless it produces all cubes—which cannot be done. The machine wears out first." (Ross, 2008, p. 198 n. 14) So the semantic version if properly understood is congenial and can be said to be a corollary of Ross's argument, but it cannot be taken as the basis for Ross's conclusion regarding the impossibility of naturalizing formal understanding. In addition, Ross's parallel argument as to content seems to hinge more on the fact that conceptual content is repeatable and shared in ways no physical thing can be and that it lacks the transcendent determinacy of the physical than on semantic determinacy *à la* Feser. If we want to properly understand Ross's insight, it is to the form of understanding that we must look first and foremost.

Feser attempts to respond to Dillard's epistemological objection, but unfortunately his reply falls flat. Feser writes,

> Dillard is essentially conflating questions about indeterminacy and questions about irreducibility, and thereby misunderstanding Ross's argument. Ross isn't arguing that

> thought is *irreducible* and therefore immaterial. Again, as an Aristotelian he would not make such an inference. Rather, he is arguing that thought has *a determinate semantic content* and is therefore immaterial. So, the mitosis/schmitosis example is simply not relevantly parallel to Ross's examples, because there is *no semantic content involved* in mitosis.
>
> (Feser, 2017; my emphasis)

But, of course, this is a *petitio*. The question is precisely why the bock should stop with semantic content, that is, concepts or the determinate content of thoughts. To answer that it should because the mitosis/schmitosis example does not involve semantic content is to beg the question. Dillard's claim is that one can concoct the same kind of thought experiment that Kripke made with respect to mathematical functions (and Goodman with respect to color expressions and Quine with respect to translated sentences, behavior, and scientific theories) using biological (or, more generally material) functions. If one can concoct the same kind of thought experiment, then it seems that the same conclusions that Feser draws using the addition/quaddition example can be drawn using the mitosis/schmitosis example. But this seems highly implausible. Furthermore, what about the concepts *mitosis* and *schmitosis*? Are these not semantically determinate? Alternatively, are our thoughts about mitosis and our thoughts about schmitosis semantically determinate? If they are, no biological process can determinately satisfy them. Otherwise they would be immaterial! This is the unfortunate consequence to which we are led if we follow Feser's semantic version of the indeterminacy argument and his response to Dillard. Since KRA does not arrive at its conclusion by means of the features of semantic content, it faces no such problems.

Now, Feser is right in pointing out that Ross's argument is not one from irreducibility. This is a confused reading of the argument, and we made our case in chapter 3. But Feser's inference from the semantic determinacy of the content or object of thought to the immateriality of thought in his response to Dillard makes it hard to see how this reply doesn't send Feser back into the hands of the content fallacy objection, unless he identifies thoughts with their contents in which case whatever is true of the content is true of the thought. But then Feser falls prey to Putnam's objection, and our thoughts of triangles end up being triangles. I am of the opinion that Feser is simply confused and does not have the 'formal' *vis-à-vis* 'semantic' distinction clear. Dillard seems more than happy to accept the semantic determinacy of thought but his explanation is that such determinacy can be due (for all we know) to some higher-level or perhaps *sui generis* physical facts. To reply that the mitosis/schmitosis example is not relevantly parallel because there is no semantic content is to beg the question, since Dillard's point is partly that it hasn't been shown that there's a categorical (metaphysical) difference between the semantic determinacy of the concept or the thought of a biological function (e.g., mitosis/schmitosis) and the semantic determinacy of the concept or the thought of a pure function. If the claim is that anything material satisfies the concepts (or the semantic contents of the thoughts) of incompossible pure functions (e.g., addition/quaddition), Dillard fires back by claiming that anything material can be shown to satisfy the concepts (or the semantic contents of the thoughts) of incompossible material functions (e.g., mitosis/schmitosis). Just as it does not follow from the latter that therefore no material function is realized at all in matter (that is, that there is no such thing as the biological process of mitosis or that, if there is, it must be immaterial!), it does not follow from the former that no pure function is realized at all in matter.[290] All that follows is, if anything, underdetermination. So Feser's response just misses the point.

---

[290] Again, 'mitosis' as a concept is semantically determinate. Indeed, Feser's response is precisely that 'mitosis' as a concept is semantically determinate but mitosis as a biological process is not (this much, properly interpreted, makes

One can see why Dillard accuses both Ross and Feser of opening up a gulf between concepts, which are semantically determinate, and the world, including behavior, which is semantically indeterminate (or underdetermined, says Dillard). How can any of our concepts be determinately of any natural, material substance and process? To be fair, though Feser does oscillates inadvertently now and then between the formal and semantic aspects of understanding, he does focus considerably his discussion on pure functions like addition and conjunction rather than natural functions like mitosis and photosynthesis. His problem is that he takes the main role of pure functions in the argument for the non-materiality of the understanding to be at the level of the variable component, that is, to be at the level of the conceptual content of thoughts (what the thinking is *about*). It is not enough to say that formal thinking is semantically determinate because (i) one can think formally about anything (physical or otherwise) in a semantically determinate way, like biological functions or juridical judgments (e.g., a syllogistic reasoning about mitochondrial DNA or about a Supreme Court ruling), (ii) formal thinking need not be about anything (e.g., greating an effectively axiomatized system with no semantics or model), and (iii) non-formal thinking can be equally semantically determinate (e.g., pondering whether to take the GRE or to make an experiment with heavy water). The solution is to abandon a general argument from the objects or content of thought, scrap semantic determinacy as the main argument for immateriality, and focus on what makes an intellectual act the kind of act it is (e.g., an act of *modus ponens*, of adding, of quadding, of conjoining, of disjoining, etc.) rather than the fact that intellectual acts can be *about* something. It makes no sense to say that mitosis is a form of reasoning, unless we make it into a norm or rule for reasoning, that is, a pure function, formal rule, or a formal specification and description. But then we've changed the subject and the meaning of 'mitosis' since it is now either a formal rule for intellectual activity or the formalization of a non-formal structure (e.g., the mathematization of a biological phenomenon). We say that this is the case with 'schmitosis', as with any other abstract functional construct that *we* can make by idealizing the process in question in numerous ways (say, by placing some numerical upper bound on the function or what have you). This is something *we do*. The crucial question for KRA is whether the forms that constitute acts of thinking or acts of understanding, like adding or quadding, and the formal theoretical specifications or constructions that are its product can be wholly physical or material. The answer is no. If a biological process is indeterminate with respect to 'mitosis' and 'schmitosis' then this means one of two things: either it means that we do not yet know what the biological process in question is and which theory is true (or truer) of it, or we do know that it satisfies determinately none in which case 'mitosis' and 'schmitosis' are idealizations, abstractions, from the real biological phenomenon. The biological cannot be biologically indeterminate. Otherwise it wouldn't be biological. Either way, formal theoretical specifications and constructions cannot be determinately satisfied, nor generated, by anything physical on account of physical properties alone. This much we have shown. So, as far as KRA is concerned, Dillard's example cannot even get off the ground.

---

sense since semantic features are linguistic-conceptual features, not biological or natural *pace* reductionists and eliminativists). But if the argument is that whatever is semantically determinate cannot be material, then the material process or structure from which the concept of 'mitosis' was abstracted cannot be said to be a determinate case of 'mitosis' as opposed to 'schmitosis'. Our concepts of material phenomena could not exclude other incompossible concepts from being the true concept of the relevant material phenomena. Now, this might be alright for an anti-realist immaterialist (I'm not even sure if they exist), but this cannot be right for Feser who holds to realism, to the formal identity (in the Aristotelian sense) between material structures and our concepts of such structures, and to the theory of abstraction (that the structures or essences of natural phenomena are abstracted by the human intellect and constituted as concepts thereof; but the general semantic determinacy argument implies that there is no fact of the matter as to whether the concept of a given material structure is the concept of *that* structure, since any material structure is semantically indeterminate by virtue of being material and will thus satisfy incompossible concepts.)

In reply to a second objection by Dillard, Feser seems to switch from an argument from semantic determinacy to an argument from formal determinacy. As is often the case, Feser seems wholly unaware of the fallacious shift:

> *Whether an L-sentence is logically valid or whether an L-sentence is a logical consequence of other L-sentences has nothing to do with whether there are determinate facts about human thinking, any more than whether ferns in the Smoky Mountains are undergoing photosynthesis has anything to do with determinate facts about Tasmanian devils. (Dillard, 2014, p. 144)*
>
> Now, Dillard's point here, as far as I can tell, is that whether an argument in such a formal language is valid or not is just an objective fact that has nothing to do with what anyone thinks about it. Hence the determinacy or indeterminacy of human thought is irrelevant.
>
> But if this is what Dillard is saying, then it seems to me that he is simply missing Ross's point. The question isn't whether there might still, as a matter of objective fact, be logical connections between propositions even if human thought was material (if we understand these objective facts in Platonic terms, say). The question is whether *human thought could ever get in contact* with these facts. And what Quinean and Kripkean indeterminacy arguments entail, Ross argues, is that human thought could not do so *if* it were material. For while there might still in that case be a fact of the matter about whether modus ponens is objectively a valid form of inference, there would be *no* fact of the matter about *whether anyone's thoughts actually conform* to modus ponens or to some other, invalid inference form instead. And *that's* the sort of result that generates the incoherence Ross is talking about.
>
> (Feser, 2017)

But what does 'conform' mean here? Is it conformity of the content of the thought or conformity of the structure of thought (i.e., the *form* of thought) to a given pure structure? Feser is right in pointing out that Ross is concerned with the fact that formal (e.g., logical) features are real features of the understanding and its manifold exercise (e.g., reasoning, calculations, inferences, and so forth.) Maybe pure forms are more than forms of understanding, but that they are and have to be at least forms of understanding is undeniable. Otherwise there would be no such thing as mathematics and logic as a human activity, and therefore no possibility of any formal science. But if Feser is using 'conformity' in terms of the content of thought, then he cannot reasonably infer from such conformity the immateriality of thought without being subject to the content fallacy and without a justification as to why these considerations do not extend to other kinds of contents (i.e., contents about material structures). If contents are identified with concepts and concepts with thoughts, then Putnam's objection arises for Feser. The semantic version of the argument can be saved and properly used only as a corollary to the argument from formal determinacy. For the forms of understanding can become the object and content of thought and in that case one can say that the content of the thought or the formal concept in question is determinate in a way no physical thing can be because we've already shown that the formal concepts in question concern operations that cannot be carried out in any intrinsic, inherent, and determinate manner by anything entirely physical. But if we begin with an argument from content, since the object of thoughts and understanding can be about non-formal structures like biological functions and processes, we run the risk of fallaciously making it so that all possible phenomena which thinking can be about are indeterminate with respect to the content of the thinking, and therefore fall prey to Dillard-like

objections and to the trap of making all physical, natural phenomena indeterminate with respect to their concepts. This is a short step away from full-blown anti-realism about concepts. Feser needs *form*—in Ross's sense of 'form'.

9. CONCLUSION: KRIPKE, ROSS, AND RULE-FOLLOWING

We have made a case for the proposition that Ross's argument is sound, and that it implies that human understanding cannot be formally indeterminate. Together with our version of Kripke's argument, we can conclude that human understanding cannot be wholly material and that KRA is sound, which is what we set out to prove. We have railed against the view that holds to the formal indeterminacy of human understanding. It amounts to sheer nonsense and cannot be coherently and intelligibly countenanced. Recall what we started with (chapters 2 and 3): Kripke's attack on functionalism/computationalism. It was directed at what he took to be functionalism's "strongest rather than its weakest link" (Kripke, 1984, p, 1). Let us restate it in the following way,

- (FBD) **Functionalism's Basic Distinction** = The basic distinction between the abstract function, description, or diagram, on the one hand, and the actual concrete physical object(s) or process that is supposed to physically realize that abstract function, description, or diagram (cf. Kripke, 1984, p. 2).

The distinction between the abstract function, description, or diagram on the one hand and the concrete physical states and processes of the physical object (brain, computer, etc.) that are supposed to *realize* the relevant abstract function, description, or diagram, is categorially correct. Unfortunately for the functionalist and computationalist, it cannot be bridged in any mind-independent way. In itself, the latter never intrinsically and determinately realizes the former.

If our intellectual activity is explained in terms of the way physical objects (in particular, physical computing mechanisms) realize, implement, or embody logical and mathematical functions, then what applies to them applies to us, and what applies to them is this: physical causal states, connections, changes, and stages alone do not suffice to determine any definite pure, abstract function. As Kripke would say, any physical computing mechanism is an imperfect realization of the abstract diagram of whatever mathematical or logical function it is supposed to 'compute'. For any physical computing mechanism taken to compute some abstract function *f* (or to realize the abstract causal diagram *F*), there will be at least some other abstract function *g* not compossible with *f* that the physical mechanism will equally well satisfy.

- Two simplified examples to illustrate Kripke's point:

  Example I:

  - Let *f* be identity function and *g* the quidentity function (say, for any *n* smaller or equal to 2, *n* itself; otherwise, *0*; cf. Chapter 1)
  - Let *h* be another function, call it the "it-function" = For any *n* smaller or equal to 2, *n* itself; for any *n* greater than 2 but smaller than 10^{856432}, *0*; otherwise $2n+1$.

We assume, of course, that '2' is either that number above which no number has been calculated or a number too great to be computable by anything with finite memory capacities, in a finite amount of steps, and in a finite amount of time. We can schematize the argument for the formal indeterminacy of the physical as follows:

(1) For any finite sequence of actually or potentially performed computations by a physical system, there will be at least two incompossible abstract functions with which all the relevant physical facts of the physical system fully accords.
(2) Yet no two incompossible abstract functions can be realized at the same time in the same respect by a given system.
(3) Thus, in order for any kind of realizer to realize an abstract function, it must realize *that and only that* function to the exclusion of any other incompossible one.
(4) Hence, since a physical system *qua physical* cannot exclude its being a case of multiple incompossible functions (say, *f*, *g* or *h*) it cannot realize any such functions.
(5) Ergo, a physical system is indeterminate among incompossible abstract functions.

The Q-D line of reasoning would have us include the 'human mind' or the 'human intellect' in the set of physical systems, objects, and processes,[291] so as to bring human understanding within the scope of (1)-(5). Ross, however, has shown this to be absurd. We can break down our 'soft' and 'hard' considerations into the following two arguments:

I. First Argument:
   (A) Human intellectual activity is a member of the set of physical objects and processes, alongside pocket calculators, computers, brains, and so on. [Assumption]
   (B) Human intellectual activity is indeterminate amongst incompossible abstract functions [By (A) and Kripke's Argument]
   (C) Therefore, human intellectual activity is indeterminate amongst incompossible logical functions or forms. [By (B) and (C), and df. abstract functions = mathematical and logical functions]
   (D) Therefore, human intellectual activity cannot realize any logical form. [From (C)]
   (E) Therefore, no human intellectual activity can be formally valid (or invalid). [From (D)]
   (F) To argue, that is, to make an argument is an instance of intellectual activity. [Def.]
   (G) Therefore, no one can argue validly (or invalidly). [by (E) and (F)]
   (H) Therefore, no one can argue validly (or invalidly) for (A).
   (I) Therefore, there cannot be any reliable inferential pathways to (A).

Note that this generalizes: there is no such thing as arguing validly or invalidly for any proposition whatsoever.[292] Two things immediately follow. It is not clear how there can be any arguing at all, since arguments are, at a suitable level of specificity, constitutively of a definite logical form. But an argument with no logical form, whether a valid or invalid one, is no argument at all. Hence, it is not possible to argue for or against anything. Logically (pun intended!), if there cannot be such a thing as arguing validly (or invalidly) for a position, then there cannot be any valid arguments for (A) and no reliable (say, truth-preserving) inferential pathways towards (A)—and for any other proposition. In

[291] In fact, this is a category mistake. For the human mind or the human intellect is neither a system, nor an object, nor a process, nor a state. It is, rather, the actualization of a capacity and therefore consists at least in part of an acquired array of abilities with respect to that which is understood. We follow Wittgenstein here, only partly.

[292] I have used 'arguing' rather than 'argument' not to beg the question against the Platonist about arguments. To the Platonist, there may still be valid and invalid arguments even if no one can argue in accord with them. So even the fact that no one can thus argue does not imply that there aren't valid and invalid arguments. Though I do not agree with the Platonist (what are arguments if not the products of arguing?), even if we concede his claim it would not change the fact that, because of (A), we simply could never grasp and instantiate in our own thinking such argument-exemplars. There would be no such thing as arguing and arguments *down here*. Everything we do intellectually would satisfy incompossible argument-exemplars at the same time and in the same respect. Hence, no such ideal arguments could be realized.

fact, there cannot be any arguments for or against (A). We are left wondering what, if any, is the rational justification for (A) and for the Q-D objection, since it looks like there can be none.

II. Second Argument:
   (J) Human 'intellectual activity' is a member of the set of physical objects and processes, alongside pocket calculators, computers, brains, and so on. [Assumption for *Reductio*]
   (K) Therefore, human intellectual activity is indeterminate amongst incompossible forms.
   (L) But to deny that we ever think in ways that are determinate in form is precisely to think in a way that is determinate in form.[293]
   (M) Hence, in order to deny that we think in a form that is determinate amongst incompossible forms we must think in a form that is determinate amongst incompossible forms.
   (N) Hence, human intellectual activity is not indeterminate amongst incompossible forms.
   (O) Hence, human intellectual activity is *not* a member of the set of physical objects and processes.

Since intellectual activity consists in the exercise of intellectual abilities, which are in turn constitutive of understanding, then those abilities whose exercise constitutes the sort of activity under consideration cannot be entirely physical nor consequent entirely upon physical capacities.

I have not said much about how formal understanding (e.g., what counts as having understood a given formal concept or operation) is a species of practice-bound understanding, embedded in normative practices (e.g., formal sciences) of a certain sort that give the criteria for what counts as having understanding of the subject of the practice. Such practices require a complex web of intersubjective elements and conditions in order any understanding, insight, and development within such practices to take place. Mathematics is an example of such a kind: a well-established set of normative practices (given the proliferation of different branches of mathematics) and a long-standing tradition with standards of correctness, criteria for excellence and competence, criteria for understanding, and so forth. The following quote points towards the right direction, but much more needs to be said:

> A rule for an arithmetical series, like an explanation of the meaning of a word, does not contain its applications *in nuce*. It is correct that acting thus-and-so counts as applying the rule correctly. But the rule is not a magical device from which the correct applications unfold. Any rule, given by a rule-formulation, can be misunderstood, interpreted wrongly, and hence misapplied. Correct application of a rule (or explanation of meaning) is a criterion of understanding. But what makes a given application of a rule correct? Not intuition; not, save trivially, accord with what the teacher meant; not, save by begging the question, doing the same as one was shown in a previous example; and not, save at the cost of a futile regress, an interpretation. In one sense, nothing makes it correct. But that is no mystery. Rather, the question is misleading. There is an internal relation between a rule and its extension. If a rule requires one to *V* in circumstances *C*, then *V*-ing in *C* is what is called 'obeying the rule'. The internal relation is forged by the existence of a practice, a regularity in applying the rule, and the normative behaviour (of justification, criticism, correction of mistakes, etc.) that surrounds the practice. Only when such complex forms of behaviour are in play does it make sense to speak of there being a rule at all, and of rule-following behaviour that accords with it. For the mere signs, the rule-formulation, are not

[293] Alternatively, to grasp and think such functions one must be able to think in a form that is determinate amongst incompossible functions.

> magic. They express a rule only if they are used as the expression of a rule, a norm of correctness, a justification for acting thus-and-so, an explanation of action, in the context of a persistent practice.
>
> (Hacker, 2005, p. 14)

A further problem that seems to lie in the way of KRA concerns the explanation of how can it be possible for formal understanding and the application of formal rules to be intersubjectively acquired, communicated, manifested, and subject to criticism if nothing physical (including behavior) can be of definite pure form. This problem, which can be made into an objection, is similar to issues concerning the knowledge of other minds, of other people's intentions and beliefs, and of meanings if meanings are taken to be "in the head". It seems that KRA opens up an unbridgeable gap between, say, mathematical understanding and the embodied behavior that is supposed to manifest and convey that understanding (Dillard, 2011, but see 2014 for a use of Wittgenstein to make the misguided claim that KRA implies that formal understanding is private). We acknowledge this worry and will consider it briefly in the conclusion as a guide to the perplexed, but a full and satisfactory treatment of this problem will have to be the object of another book.

In addition, we have not considered some of the implications that KRA has for other philosophical domains and projects, including philosophical presuppositions about the nature of language in linguistics, the nature of physical phenomena in relation to mathematics in physics and, more broadly, the applicability and explanatory role of formal structures (e.g., mathematics) to matter, etc. It is impossible to consider in depth all of these issues, nor the wider implications our argument has for the philosophy of cognitive science, for philosophy in general, and for the proper explanation of the possibility of disciplines who require the application of formal understanding (e.g., engineering). But it seems fitting to briefly explore, albeit neither exhaustively nor comprehensively, those aspects of our argument that may have some bearing on the nature of those abilities and activities which are either in whole or in part constituted by (or necessary for the acquisition and) exercise of formal understanding. We shall consider our capacity for linguistic understanding. This is the topic of the next, concluding chapter.

6
KRA, Language, and Linguistic Understanding: Some Considerations

In this final chapter, I would like to extend some of the previous considerations to the philosophical exploration of the nature of language and to certain linguistic theories, in particular Chomsky's generative grammars. Although KRA can be said to have implications for *any* formal approach to our capacity for language and to linguistic understanding, I would like to explore the consequences it has for Chomsky's approach in particular. One of the reasons for this is that Chomsky is commonly regarded (see e.g. the papers in McGilvray, 2005) as having inaugurated the naturalistic approach to language, that is, as having set the study or the "science" of language (linguistics) on the path of a natural science, using formal (mathematical) theoretical constructions analogous to those of physics and chemistry. For Chomsky, linguistics is, ultimately, a branch of biology. But Chomsky posits a universal grammar that is innate to our language capacity and which is a formally definite structure wholly determined (or predetermined) by our species-specific biological endowment and realized in an organ ("the language organ") or a component of the human brain.[294] KRA if sound shows that this proposition cannot be intelligibly and coherently sustained.[295] Another reason, not unrelated to the first, is that Chomsky himself has speculated that (the capacity for) mathematical and logical understanding might have developed as a by-product of the language faculty (see e.g. Chomsky, 1980, p. 249) and is essentially an "abstraction" from the mechanism of discreet infinity that is unique to human language (see e.g. Chomsky, 1988, p. 169). The problem we wish to address is not nativism about language but rather Chomsky's categorical characterization of the innate properties or features of the language faculty. It is practically impossible to consider here in detail Chomsky's work on language which spans over more than five decades. Nor is it practically feasible to consider in detail all of the implications KRA has for assessing the criteria of intelligibility that any theoretical framework that purports to account for the capacity for language and for linguistic understanding must meet. Our aim here will be merely to sketch and explore the philosophical significance of KRA relative to what I call Chomsky's conjunction:

C1. The language capacity ("the language faculty") has a determinate innate formal structure, *and*
C2. The language capacity or faculty is wholly neurobiologically determined and realized.

The second conjunct, namely C2 captures Chomsky's commitment to what I call 'neurobiologism': that the capacity for linguistic understanding is *wholly* a matter of the species-specific neurobiological endowment of human beings. Given the plausible assumption that the language faculty (i.e. the capacity for language) is innate to human beings together with the more contentious C1, it follows that the innate formal structure of the language faculty is wholly neurobiologically determined.

---

[294] "It is hard to avoid the conclusion that a part of the human biological endowment is a specialized 'language organ,' the faculty of language (FL). Its initial state is an expression of the genes, comparable to the initial state of the visual system, and it appears to be a common human possession to a close approximation. […] we can think of a particular language L as a state of FL. L is a recursive procedure that generates an infinite of expressions." (Chomsky, 2002, pp. 85, 86–87). To wit: "It could be, then, that *the recursive computational system of the language organ* is fixed and determinate, an expression of the genes, along with the basic structures of possible lexical items. (Chomsky, 2002, p. 89)

[295] Just as we've rejected individualism but not the individual in an explanation of formal understanding (see chapter 1), we likewise reject neurobiologism but not neurobiology in an explanation of the capacity for language. These distinctions must be kept in mind.

Neurobiologism is a response to both the constitutive and genetic problem (see chapter 1) of the nature of the language capacity, which is the capacity for linguistic understanding. If that capacity is formally determinate, as it surely is for Chomsky,[296] or formally determinable (or both) then it cannot be determined nor explicable wholly materially, whether in neurobiological or other purely material terms. Thus, linguistics cannot be, wholly, a branch of biology nor, wholly, a naturalistic inquiry.[297] A corollary of our objections to Chomsky is the observation that his dismissal of the mind-body problem as unformulable given the absence of a definite and coherent concept of body (Chomsky, 1988, 1995a, 2002) is not entirely to the point. For all it's worth, he might correct about that and we have not the intent to weigh in on the cogency or lack thereof of the so-called mind-body problem. But we will point out that Chomsky is nevertheless subject to, and quite actively engaged in, a variant of the mind-body problem: call it 'the mind-brain problem'. Since there *is* a coherent and definite concept of the brain (always completable and rectifiable, of course, given our limited understanding of the human brain)—or, at least, since Chomsky takes it for granted that there is such a thing as a coherent conception of the brain—and given that he himself constantly refers to the language faculty as a component of the human brain, the problem as to whether the sort of formal structures and operations which Chomsky attributes to the language faculty can be intelligibly attributed to the brain, or to components of the brain (i.e. "the language organ"), arises all the same.

A clarification. It is usually said that the language faculty, contrary to a natural language (what Chomsky calls an I-language, see §1 below), is *structure-independent*: exposure to a particular language under suitable conditions generates a structure, that is, a particular grammar in the individual that becomes the particular language the individual knows. But the language faculty, though structure-independent in this sense (that is, in the sense that it lacks in its initial state the particular structure of any particular grammar), is not structureless: it is a complex system, formally and semantically rich in abstract *constraints* (e.g. rules, principles and parameters, recursive operations, etc.) that pre-fix the possible particular grammars that may be generated and thus delimit the realm of possible human languages that any individual human being can acquire. The language faculty is not a *tabula rasa* or a blank slate (Pinker, 2003) but rather a genetically determined computational system whose abstract formal structure enables and plays a role in the acquisition and constitution of any particular language and which is to be expressed and captured by theories of generative grammar. This innate structure of the language faculty is what Chomsky calls universal grammar (UG, for short).

But this gives rise to innumerable problems. Is the language faculty as formally rich and determinate as the (correct) formal theory proposed by the theoretical linguist would have it be? If

---

[296] "…[Universal Grammar] is a theory of the 'initial state' of the language faculty, prior to any linguistic experience." (Chomsky, 1986, pp. 4–5)

[297] An interesting but difficult topic that we have neither the time nor the space to consider in this monograph is the explanatory role and the applicability of formal (mathematical) constructions in the natural sciences, in particular in physics and chemistry. Perhaps someone might be tempted to think that our argument, KRA, leads to an anti-realism about the use of mathematics in the study and sciences of natural phenomena. But this is not so. So long as the formal (mathematical) construction of physical and chemical phenomena are not identified with the physical and chemical phenomena being formalized (mathematized), KRA has nothing to say about the use, the (explanatory) value, and the applicability of formal constructions or tools to the study of matter. Likewise, granted the aforementioned proviso, KRA says nothing about the use of material structures and mechanisms (many artificially constructed) to simulate or physically represent formal structures. We hold that both cases are indeed possible if properly understood, but the onus is on us to explain how such a thing could be possible and what would the significance of it be. Yet this would lead us into the philosophy of mathematics and philosophy of science, all of which is well beyond the scope of the present work. The problems are, however, admittedly very important and quite complex. For some possible accounts compatible with our line of reasoning, see e.g. (Cartwright, 1983, 1989, 1994, 1999, 2007; Shanks and Brzeziński, 1998; Niiniluoto, 2002)

Chomsky were to reply by insisting that only particular grammars are formally structured such that formal determination takes place only at the level of generated, particular languages and not at the level of the language faculty, then his neurobiologism about the language faculty will be unaffected by KRA. But then the language faculty, on account of KRA, cannot be a formally determinable thing since it is a wholly material faculty. Hence, the language faculty could not be used to explain, for it could not explain, the generation of formally determinate structures (i.e. particular grammars). But this is precisely what the language faculty was supposed to explain. If, however, Chomsky doubles down on what I take to be his position: that the language faculty is *both* formally *determinate*, innately so in virtue of definite abstract constraints (e.g. principles and parameters), and formally *determinable*, through exposure to particular grammars in the form of linguistic data within a particular linguistic community, then given KRA his neurobiologism about the language faculty must be abandoned on pain of contradiction. For no material thing can be either formally determinate or formally determinable in any intrinsic and non-relative (say, observer-dependent) way *on account of material properties alone*. If, however, Chomsky were to reply instead with the statement that formal determination takes place only at the level of the linguist's theories and models rather than at the level of that which the theory aims to characterize and represent, then language and the language faculty may still be naturalizable but at the cost of turning UG theories into instances of formal (e.g. computational) *modeling* rather than formal (e.g., computational) *explanations*, for formal models (e.g. computational models) characterize complex systems by means of formal structures (e.g. computations, algorithms) without thereby necessarily attributing any such structures to the systems that they model (consider for example a computational model of the weather).[298] This is all well and good except for the fact that Chomsky's project for linguistics now loses its *explanatory* and *realist* import insofar as the *computational* mechanisms that figure in, and are posited by, the theoretical constructions of universal grammar turn out not to be *real and intrinsic constituents* of (the universal grammar of) the language faculty; what's more, it would now remain that the linguist's capacity and ability to carry out such a formally determinate intellectual undertaking would not be naturalizable, regardless of whether the language faculty is. Plus, the formal-theoretical constructions and models of the linguist would be formal idealizations and abstractions that would thus be imperfectly realized or approximated by the physical mechanisms they purport to represent.

---

[298] Recall our discussion on Piccinini's mechanistic account of physical computation and the difference between models and explanations in chapter 3. The two are not equivalent. To provide a model of some phenomena is not, in and of itself, to explain the phenomena. "Someone might insist that … the goal of explanation is to capture in a model how a system behaves; models need not describe components in order to capture how a system behaves, … in order to explain. The problem with this view… is that it confuses modeling with explaining. A resounding lesson of 50 years of sustained discussion of the nature of scientific explanation is that not all phenomenally and predictively adequate models are explanations. We can construct models that predict phenomena on the basis of their correlations (barometers predict but do not explain storms), regular temporal successions (national anthems precede but do not explain kickoffs), and effects (fevers predict but do not explain infections). Furthermore, there is a fundamental distinction between redescribing a phenomenon (even in law-like terms) and explaining the phenomenon. Snell's law predicts how light will bend as it passes from one medium to another, but it does not explain why light bends as it does. We might explain that the light bent because it passed from one medium to another, of course. But that is an etiological explanation of some light-bending events, not a constitutive explanation of why light bends when it passes between different media." (Piccinini, 2015, p. 94) Models may be descriptively adequate or inadequate, but issues of explanatory adequacy are constrained by considerations of explanatory relevance, in particular the behavior or function of components of the system or mechanism whose system-level behavior is being modeled. Otherwise, as Piccinini points out, it will not be "possible to distinguish explanations from merely predictively adequate models and phenomenal descriptions of the system's behavior." (Piccinini, 2015, p. 94) Chomsky is sensitive to the difference between descriptive and explanatory adequacy (see e.g. Chomsky, 2002, p. 93). This is partly why our argument is of consequence for Chomsky's conjunction, because a constraint placed by him on the explanatory adequacy of any UG theory is the success in explaining the computational mechanisms of the language faculty, an organ of the human brain/nervous system.

The issues surrounding the indeterminacy of physical mechanisms *vis-à-vis* their formal (-mathematical) representations and specifications discussed in chapters 2 and 3 would now plague the linguist's theories since any physical mechanism (including the brain) that is supposed to count as the language faculty (or any physical mechanism of which the language faculty is supposed to be a component) will be indeterminate with respect to any formal structure and operation including those posited by Chomsky's theories of universal grammar. Incompossible grammars (say, qu-like morpho-syntactic generative rules or qu-like principles and parameters) would be equally satisfied by the putative physical mechanism. On the other hand, if the linguist's formal-theoretical models *do* indeed purport to model and characterize the real and inherent formal structure of the language faculty such that what is being modelled is indeed the formally determinate structure of the language faculty, then we are back at square one: Chomsky's neurobiologism about the language faculty cannot be intelligibly and coherently sustained. This is the crux of our argument against Chomsky's conjunction.

Due to the nature of the issues involved and the limits and scope of our investigation, our treatment of these issues will be selective and limited rather than exhaustive and complete and will consist more of an exposition of problems than a full-blown refutation or solution to the difficult topics and problems surrounding the nature of natural language and our capacity to acquire and understand languages.[299] This chapter is organized in four main parts. In the first part we give a brief characterization of Chomsky's cognitivist or Cartesian approach to linguistics (what we here call 'theoretical linguistics'), the distinction between E-languages and I-language approaches to language, his preference for the latter over (and, at times, downright dismissal of) the former[300], and how an I-language is made possible in the individual partly by an innately formally determinate faculty (i.e. a computational system[301] of some sort, variously characterized through the history of generative grammars depending on the prevailing model). In the second part we consider the major implication that our argument, KRA, has for Chomsky's conjunction: either Chomsky accepts Universal Grammar (UG) about the language faculty or he accepts neurobiologism about the language faculty, but not both.[302] In the third part, we consider Chomsky's own criticisms of Kripkestein (Kripke, 1982) and the rule-following considerations therein. We find them wanting and misplaced and apply KRA to the philosophical underpinning of Chomsky's approach to linguistics, in particular the constitutive and genetic explanations of the language faculty and linguistic understanding. We end

---

[299] The following summary of Chomsky's views relies heavily on (Cook and Newson, 2007) but all relevant primary sources have been consulted and quoted directly from.

[300] "E-language, if it exists at all, is derivative, remote from mechanisms and of no particular empirical significance, perhaps none at all." (Kasher and Chomsky, 1992, p. 10) An E-language consists in the set of actually produced utterances and sentences by the members of some linguistic community. Approaches to language that take language to be essentially a social phenomenon and that consider natural languages are wholly constituted by practices, use, culture, history, evolution (e.g. optimal communication), etc., are misleadingly lumped together under the (for Chomsky, pejorative) label of E-approaches to language and as dealing with E-languages. Naturally, Wittgenstein is considered by Chomsky as an exemplary case of a philosophical E-approach to language.

[301] It should be noted that computational linguistics is not *ipso facto* I-linguistics. Computational analysis can be used by both E- and I-approaches to language, but the former would be concerned with the analysis and synthesis of a large amount of data taken from actual speech and linguistic behavior whereas the latter would be concerned with the formulation of the rules or computational operations that are innate or internal to the mind of the speaker.

[302] Note that we will, for reasons of simplicity, stick to talking about 'faculty' or 'capacity'. But as our brief exposition of our argument against Chomsky's conjunction in the introduction to this chapter should've made clear, the same considerations can be applied to our concept of 'ability' or 'linguistic understanding' (in Chomskyan terms, to the particular grammar the speaker-hearer knows and understands after going from the universal initial state S_0 to the particular steady, stable state S_L that constitutes the speakers' understanding of the particular grammar or language, e.g., her native tongue) if linguistic understanding is held to be formally determinate, as Chomsky surely does.

with some brief comments on the nature of language, the capacity for language, of linguistic understanding and activity.

1. CHOMSKY ON THE NATURE OF LANGUAGE, THE LANGUAGE FACULTY, AND LINGUISTIC UNDERSTANDING

According to Chomsky (Chomsky, 1995b) the aims of linguistic are threefold, with a fourth one sometimes added (Chomsky, 1988, p. 3, 1986), and can be formulated in order of conceptual importance as follows:

(i) What constitutes knowledge of language?
(ii) How is such knowledge acquired?
(iii) How is such knowledge put to use?
(iv) What are the physical mechanisms that serve as the material basis for this system of knowledge and for the use of this knowledge?

The 'system' referred to in (iv) is computational[303], so note that (iv) can already be considered a loaded question if it presupposes the determinate physical realization of a computational structure. As Chomsky himself acknowledges, any answer to (ii) and (iii) will depend on the answer given to (i), and the same goes for (iv). Our strategy is not to argue against Chomsky's answer to (i) *simpliciter*. We leave it open whether mastery of a language is constituted and made possible by an innate computational structure (specifically, form the 2000's onward, an innate structure with built-in computational mechanisms for recursion; see Hauser et al., 2002) whose 'triggering' or 'activation' on account of exposure to linguistic data and experience 'triggers' or 'occasions' the generation of a particular grammar (that is, a specific language like English or Chinese) in the mind of the individual (what Chomsky calls *competence*, mastery of the grammatical structure and rules of a language). Issues about innateness are not our concern here, and neither are issues surrounding the question of whether language learning and mastery involves computational operations at all. Likewise, we have little to say about Chomskyan answers to and investigations into (ii) and (iii). Our aim will be rather to lay out the difficulties that lie in the way of Chomsky's conjunction as manifested in his answer to (i) and to (iv). Cook and Newson, following Chomsky, explain: "[t]his mental knowledge [knowledge of a language] must have some physical correlate, in other words there must be some relationship between the computational system in the mind and the physical structures of the brain. The principle and parameters of UG [Universal Grammar] must be *stored* somewhere within the speaker's *brain*." (Cook and Newson, 2007, p. 12, my emphasis) (Notice the fallacious shift from *correlation* to *storage*.) As KRA shows, any formal structure can only be stored or realized in something material in an observe-dependent and relative way,[304] which is precisely the contradictory of the manner in which the innate computational system of the language faculty posited by Chomsky is said to be in the brain. The fact that there is no material explanation of the formal component of the mind/brain is not, as Chomsky seems to think, a matter of current practical limitations on our scientific

[303] For reasons of space and simplicity, I omit the computational mechanism attributed to the brain by Chomsky responsible for relating what Chomskyans call Phonetic Form (PF) and Logical Form (LF). If the reader so desires, she can rehash our argument in terms of the computational mechanism involved in the interaction between PF and LF. We shall focus for the most part on the generative aspect of Chomsky's approach to language as it pertains the (*in principle*, potentially infinite) production and comprehension of syntactically well-formed structures, the conditions for the possibility of this sort of capacity and ability, and the constitutive and genetic explanation given by Chomsky of such a capacity and ability.

[304] Can there be conventions of the relevant sort without language, linguistic practices, and linguistic understanding?

advancements, as if the reason were that "current understanding falls well short of laying the basis for the unification of the sciences of the brain and higher mental faculties, language among them" (Chomsky, 2005, p. 143) and therefore though "a product of biological evolution" the ways in which it is so "lie well beyond current [scientific] understanding." (Chomsky, 2002, p. 46) The limitation is, rather, an *in principle* (categorical) limitation that undermines the naturalistic presuppositions with which Chomsky constantly pumps theoretical linguistics.

1.1 The Central Project of Theoretical (Chomskyan) Linguistics

The Central Project of theoretical linguistics (a label we borrow from Wright, 1989 to characterize Chomsky's delimitation of the central or most fundamental issues for linguistics) consists in achieving an understanding and explanation of at least two crucial aspects of language and linguistic understanding: what Chomsky has called the *creative* aspects of language (the creativity problem) and the *stimulus-free* aspects of language (the poverty of stimulus argument). At times Chomsky speaks as if the former were *the* central aspect, but both are taken to be fundamental.[305] Simplifying, we can say that these aspects of language encompass (1) the language knower's ability to generate by finite means a potential infinity of sentences and grammatical iterations, (2) the language knower's ability to generate, understand, and judge correctly the well-formedness of an indefinite number of sentences never before generated and encountered in her linguistic experience, and (3) the language knower's ability to produce any words and grammatical structures irrespective of any environmental stimulus, depending on her state of mind. This last point brings us to the historical unboundedness of language knowledge and what has come to be known as the 'logical problem of the acquisition of language' (see e.g. Cowie, 1997): how can children go from apparently having no knowledge of a language to knowing (that is, mastering) a language in an utterly spontaneous way, effortlessly and rapidly, on the basis of a very limited number or sample of linguistic data, and without any explicit, overt training? The idea is that the knowledge attained when the child or a person masters a particular language far exceeds the linguistic data to which the individual could have been exposed during the acquisition of the language. What has come to be known as the poverty of stimulus argument amounts to, roughly, the claim that the primary linguistic data (*pld*, for short) to which the child is exposed (e.g., parents' language use and corrections, the conditioning and reinforcement of surrounding language speakers, etc.) during language learning is much too impoverished and small to give rise to the wealth of pragmatic, semantic, and syntactic knowledge that mastery of a language involves. In other words, the *pld* are insufficient to enable the acquisition of the vastly complex rule-system or principle-parameter system that constitute the grammar of the particular language attained (viz., the *pld* are grammatically *underdetermined*).

As Fiona Cowie explains,

---

[305] "The central fact to which any significant linguistic theory must address itself is this: a mature speaker can produce a new sentence of his language on the appropriate occasion, and other speakers can understand it immediately, though it is equally new to them. Most of our linguistic experience, both as speakers and hearers, is with new sentences; once we have mastered a language, the class of sentences with which we can operate fluently and without difficulty or hesitation is so vast that for all practical purposes (and, obviously, for all theoretical purposes), we may regard it as infinite. Normal mastery of a language involves not only the ability to understand immediately an indefinite number of entirely new sentences, but also the ability to identify deviant sentences and, on occasion, to impose an interpretation on them." (Chomsky, 1970a, p. 7) "An interest in the nature and origins of the 'marvelous invention' [the creative aspect of language] leads to an investigation of the component of the human brain that is responsible for these unique and indeed wondrous achievements [like the capacity of a language speaker-hearer to generate and comprehend a potential infinity of grammatically correct expressions on the basis of finite means]." (Chomsky, 2002, p. 47)

> … many theorists have recognized the crucial part played by language in our lives, and have speculated about the (syntactic and/or semantic) properties of language that enable it to play that role. However, few had much to say about the *properties of us* in virtue of which we can learn and use a natural language. To the extent that philosophers before the 20th century dealt with language acquisition at all, they tended to see it as a product of our general ability to reason — an ability that makes us special, and that sets us apart from other animals, but that is not tailored for language learning in particular… All this changed in the early twentieth century... With advances in syntax and semantics came the realization that knowing a language was not merely a matter of associating words with concepts. It also crucially involves knowledge of *how to put words together,* for it's typically sentences that we use to express our thoughts, not words in isolation. If that's the case, though, language mastery can be no simple matter. Modern linguistic theories have shown that human languages are vastly complex objects. The syntactic rules governing sentence formation and the semantic rules governing the assignment of meanings to sentences and phrases are immensely complicated, yet language users apparently apply them hundreds or thousands of times a day, quite effortlessly and unconsciously. But if knowing a language is a matter of knowing all these obscure rules, then acquiring a language emerges as the monumental task of *learning* them all. Thus arose the question that has driven much of modern linguistic theory: How could mere children learn the myriad intricate rules that govern linguistic expression and comprehension in their language — and learn them solely from exposure to the language spoken around them?
>
> (Cowie, 2008, pp. 1–2, cf. 1999)

Chomsky's answer is that they don't. That is, they don't learn them at all—the fundamental structures that allow for the acquisition and mastery of a particular language are already there in the child's mind/brain and known, albeit unconsciously. They are innate or inborn, part of the initial state of the language faculty with which we are biologically endowed. This is the nativist aspect of Chomsky's Central Project of theoretical linguistics. But the feature of the Central Project most relevant to us lies not in Chomsky's nativism but in his specification and characterization of that which is innate, that which enables the acquisition and generation of particular languages: the structure (i.e. the rules or the principle and parameters and mechanisms) of the language faculty.

1.2 Competence and Performance

Chomsky distinguishes between externalized language (E-language) and internalized language (I-language) which gives way to the distinction between E-language linguistics and I-language linguistics (see e.g. Chomsky, 1986, p. 20). Roughly, E-language linguistics is a statistical and social approach to language based on actual linguistic behavior and activity. It is an approach that collects data from actual linguistic activity (e.g. samples of sentences, speech, etc.) and on the basis of this proceeds to describe the properties of the relevant language and construct a grammar on the basis of the regularities found in the data or samples independently of any "properties" of the mind (for one contemporary example of this approach, see Hunston and Francis, 2000). E-language approaches usually take language to be essentially a social phenomenon and its study is a study of the actual structures or patterns that are regularly found in linguistic activity and behavior (Chomsky, 1988, p. 36). One could say, roughly, that E-language approaches are concerned with what Chomsky calls *performance*, that is, the actual use or usage of the knowledge of a language by competent speakers and therefore language as exhibited in and dependent upon particular situations, contexts, intentions, and so forth. Performance, in short, refers to the production of actual utterances of a language. But

what accounts for performance, in particular grammatically correct performance and the ability to discern its grammatical correctness or incorrectness?[306] Chomsky answers with the concept of *competence*, which is an internal state of the mind/brain of the language knower and is the final state attained through the interplay of the innate structure of the language faculty and its' exposure to a given linguistic experience. Competence is the *inner* steady state attained in the knowledge of the grammar of a particular language (say, knowledge of what *can* and *could* be done in the language) and is, to a great extent, independent of *performance* (that is, the concrete, *actual* use of the language).

An I-language, or more precisely *knowledge* of an I-language, constitutes competence (Chomsky, 1965). Therefore, I-linguistics, in contrast with E-linguistics, is concerned with discovering the 'internal' properties of language knowledge, in particular the properties that constitute a speaker's knowledge of a language as well as how and whence it comes about—problems which the I-linguist says cannot be properly addressed at the level of performance and require investigating the internal structure or properties of the mind/brain of the competent language speaker. "By 'grammatical competence' I mean the cognitive state that encompasses all those aspects of form and meaning and their relation, including underlying structures that enter into that relation, which are properly assigned to the specific subsystem of the human mind that relates representations of form and meaning." (Chomsky, 1980, p. 59) Competence, in short, refers to that which accounts for the grammatically correct production and comprehension of actual and *potential* (included never encountered, new, and potentially infinite number of) utterances and for what the speaker-hearer knows *can* and *could* be done in the language. I-approaches to language (e.g. the English competence of a native English speaker) focus on what the speak-hearer knows *can* and *could* be done in the language, rather than on what is *actually* done. The structure of grammatical competence explains why the sentence

(1) Why are you making such a noise?

is possible in English whereas the sentence

(2) *Why you are making such a noise?[307]

is not (Cook and Newson, 2007, p. 16). At the level of performance, both may convey the same meaning, say, due to the context but a competent speaker would be able to recognize that (2) is ungrammatical though contextually rendered meaningful (e.g. one can imagine a conversation between a native English speaker and a non-native English speaker in which the latter utters (2) and

---

[306] I am relegating the concept of performance to E-approaches, though this is somewhat misleading. I-approaches have much to say about *grammatically correct* performance but mostly to the extent that such performance manifests linguistic competence. The E-language vs. I-language divide does not align exactly with the performance *vis-a-vis* competence distinction (see e.g. Chomsky, 1965, p. 3). Grammatically (as opposed to contextually, institutionally, or situationally) correct performance is, for Chomsky, only properly explicable in terms of competence, which is situation-independent and enjoys a level of abstraction and universality not enjoyed by performance. Plus, performance is affected by innumerable factors (e.g. tiredness, sloppiness, distraction, time, etc.) which do not affect nor hold for competence. Competence is, in every explanatorily relevant order, *prior* to performance. Linguistic understanding then, we are told, must be couched in terms of the concept of competence, the proper subject of I-approaches. Having said that, I will talk as if performance were the mere subject matter of E-approaches. The distinction between E-language and I-language has been contentious and controversial from the start, and the same can be said for the performance and competence distinction. We are not here to settle these issues.

[307] In linguistic text books, an asterisk usually signifies an ungrammatical sentence. We follow that convention here.

is understood by the former even though it is recognized by the native speaker as ungrammatical).[308] There is a structure and form to the grammatical sentence which the competent speaker knows and which allows her to comprehend, recognize, and generate ever new sentences with the same form and structure, independently of how the sentence is used. This is grammatical and linguistic competence, and it abstracts away from praxis and use.[309] It is also the source of 'linguistic creativity' in the Chomskyan sense of 'creativity': the ability to understand and generate an indefinite number of sentences that are not part of one's linguistic experience and are thus *novel* (never before encountered or produced) yet known to be correct. Another example given by Cook and Newson is the following sentence,

(3) Ornette Coleman's playing was quite sensational.

which the competent English speaker will understand and know to be grammatically correct even if she has no idea who Ornette Coleman is and even if she has never encountered this sentence before. Likewise, she will know that the sentence

(4) *Is John is the man who tall?

is not a possible sentence of the language (that is, not a grammatically correct sentence, a sentence of a grammatically incorrect form) even if she has never heard or encountered this sentence before.[310]

Now, what exactly is the nature of grammatical competence?[311] In one sense, competence is quite congenial to our use of the concepts of 'understanding' and 'abilities'. In another, it is quite antithetical insofar as it is made possible by a determinate formal and semantic structure physically realized that precedes it (innate to the brain of the human person) and which makes competence

---

[308] This example is imperfect, but it does not affect out point. The imagined scenario falls more on the side of what Chomsky calls "pragmatic competence", which is linguistic knowledge situated in certain institutional and cultural settings, like knowing that 'what's up?' it's a greeting equivalent to 'how are you doing?' or 'how's everything going' rather than a question about what lies above you. Pragmatic competence cannot be identified with grammatical competence since one can have the latter without having the former (e.g. consider: looking up when someone greets you with 'what's up?'; that manifests grammatical competence but not pragmatic competence). We won't discuss pragmatic competence any further as it is immaterial to our discussion. For more on pragmatic competence, see (Chomsky, 1980).

[309] There is a clear derision of sociolinguistics on the part of Chomsky and of any other theory that takes communication (a social phenomenon that requires intersubjectivity) to be an essential and constitutive feature of language. Chomsky has even suggested that communication might turn out to be an epiphenomenon of language (see e.g. Chomsky, 2002, p. 107), which has as its primary and essential role the expression of thought (even if there were no other people).

[310] Creativity is a big thing in Chomskyan linguistics. It is constrained since it is not random and chaotic; yet within the grammatically relevant constraints, creativity is potentially infinite. Another feature of Chomsky's approach to language is it's strict separation between mastery and use and syntax and semantics. The now famous sentence

A. Colorless green ideas sleep furiously

was given by Chomsky (Chomsky, 1957) to showcase that syntax is separate from semantics. For although A makes no sense semantically speaking, it is a syntactically well-formed expression and therefore an instance of a grammatically correct form. It's structure can be captured or represented formally in various ways, from the rewrite rules for phrase structures (see 2) of the early days of generative grammar, later strengthened under the X-bar theory, to the most recent minimalist program and its conception of the computational mechanism of the language faculty. The normativity of grammatical forms is still taken to be determined irrespective of the intelligibility of the content.

[311] Chomsky's competence-performance dichotomy is controversial and, to the estimation of some, now widely unpopular amongst psycholinguists. I suppose this is so, in part, because it makes performance (i.e., the use of language) accidental and epiphenomenal to linguistic competence (i.e., the knowledge of language).

possible with respect to any possible natural language.[312] So universal grammar, as the content or structure of the language faculty, is that which makes possible (and determines what is possible in) the generation and attainment of any I-language. As Chomsky aptly explains, "… take the I-language to be a rule system of some sort, a specific realization of the options permitted by UG, fixed by presented experience." (Chomsky, 1986, p. 46) So, since linguistic competence is the particular grammatical 'rule' system that results in the individual from the complex combination of linguistic input in presented experience and the innate *constraints* (e.g. the principles and parameters) of the language faculty, a subset of whose settings or options are triggered by the presented experience of the individual, the relevant activated settings of the language faculty that made possible and determine the generation of a specific, particular language (I-language) must figure in the competence and knowledge of the language of the speaker. The study of grammatical competence (for example, through the study of the judgments of grammatically competent individuals) should therefore reveal the formal structures and operations (e.g. the principles and parameters, the computational mechanisms) that are *universal* and *innate* as opposed to those that are local and language-specific (that is, specific or peculiar to a particular natural language). Thus, the study of I-languages should lead to the study and discovery of the innate structure (the universal grammar) of the language faculty.

1.3 Competence and The Language Faculty

Consider the notion of Chomskyan Universals (call them 'UG universals').[313] Following Chomsky, we can say that an I-language denotes a psychological part of individual persons that is internal to them. When a person acquires some language L, their language faculty has been set to the cognitive state we can denote as 'S_L', which is the I-Language L. This comes about through a complex process of interplay between the formal rules or principles and parameters innate to the

---

[312] Grammatical competence is also, for Chomsky, tacit or unconsciously held. So competent speakers know a particular grammar without knowing, except tacitly, the grammar itself (otherwise called 'cognizing'). This is another point of disagreement between us. Chomsky compares mastering a language to 'growth' rather than to 'learning', more like the way a seed grows into a plant than the way in which a skill is acquired or the way in which facts are learned. "In certain fundamental respects we do not really learn language; rather grammar grows in the mind." (Chomsky, 1980, p. 134) It is all there from the beginning in potency, with experience being the trigger that allows for the actualization of the potentialities (like the environment for the seed). Note that although this may sound Aristotelian, it is not. What is said to be there in potency is an actually well-structured formal system of principles and parameters that only need to be 'triggered' or set. There is no learning a language, only language growing in the human person as she is exposed to a suitable (linguistic) environment. Variations in languages in individuals is explained by exposure to different linguistic environments, but the system (the seed, so to speak) is the same or of the very same nature and structure in all human persons. But see note 321 for a caveat. It is important to bear in mind that Chomsky has an idiosyncratic (and some (see e.g. Hacker, 1990) would claim nonsensical) use of pretty much every concept interconnected with the concept of language, including the concept of language itself. That is to say, the concepts of knowledge, understanding, learning, rules and following rules, mind, and pretty much every psychological and cognitive concept.

[313] Chomsky grants that the language faculty or faculty of language can be considered narrowly or broadly. Thus, the distinction between the Faculty of Language in the Narrow Sense (FLN) and the Faculty of Language in the Broad Sense (FLB). The FLN is characterized by a computational mechanism for recursion (Hauser et al., 2002) Our argument does not depend on this differentiation; it cuts through both. Note that there are many aspects of language that do not fall within the theoretical purview of UG theory. This is cashed out in terms of the distinction between core and periphery. The core of a grammar or of grammatical competence is that which is predictable and consequent upon UG; it is where all the UG principles are found and maintained. The periphery is all of that which is accidental to grammar and thus not predictable from UG. "What a particular person has in the mind/brain is a kind of artefact resulting from the interplay of accidental features… a core language is a system determined by fixing values for the parameters of UG, and the periphery is whatever is added to the system actually represented in the mind/brain of a speaker-hearer" (Chomsky, 1986, p. 147).

language faculty and the linguistic experience and community L to which the person is exposed. Let H stand for human person, and L for a particular language or grammar (e.g. English),

> For H to know L is for H to have a certain I-language. The statements of the grammar are statements of the theory of mind about the I-language, hence structures of the brain formulated at a certain level of abstraction from mechanisms. These structures are specific things in the world, with their properties… The I-language L may be the one used by a speaker but not the I-language L' even if the two generate the same class of expressions (or other formal objects) … L' may not even be a possible human I-language, one attainable by the language faculty.
>
> (Chomsky, 1986, p. 23)

Notice that the structures are structures of the brain but formulated at certain level of abstraction in terms of formal (e.g. computational) structures. The language faculty (also referred to by Chomsky as 'the language organ') comes with a built-in computational structure that determines the possible states and values (parameters) that it can be in or that it can take so as to be propelled into a definite I-language state. Thus, when someone knows a language L, she has a certain I-language, namely L, and one's language faculty is in state I-language$_L$ or S_L. When the I-linguist studies what a language speaker$_L$ knows about L she is after whatever is in the mind of the speaker$_L$ who knows L, and this is so for any language someone knows. In addition, she is trying to uncover or discover the structure of UG by means of trying to locate universal principles that play a role in a given I-language. For example, suppose L is the English language. Take the Principle of Locality or Locality Conditions, postulated by UG theory as a universal. The principle of Locality states that movements must be local, that is, short. What does this mean? Consider English questions, which involve what are called **wh-elements** (i.e. *who*, *where*, *what*, and so on) and which may be formed by moving a wh-element to the front of the sentence (Cook and Newson, 2007, pp. 20–23; 32–35), such as when the sentence

(5) Mike knows Bert.

can be transformed into a question by placing a wh-element in front of the sentence

(6) Who does Mike know?

In (6) the wh-element is the object of the verb, but since the object of a verb usually follows the verb in English, theoretical linguists say that in (6) the object of the verb has *moved* from its usual position in order to form the relevant question. Question-movement is not a principle, however; it is a parameter of variation, as not all languages (e.g. Bahasa Malaysia, Japanese) use or require question-movement. But that movements must be local if they occur *is* a universal since, or so goes the hypothesis, wherever there is movement it must be the shortest possible.[314]

Consider another explanation in terms of parameter-settings (see e.g. Chomsky, 1993a). Take the fact that all languages require that sentences have a subject, there are variations between different languages with respect to how the subject can be expressed grammatically in a sentence. In Spanish,

---

[314] For what it means for movement to be local, see the explanations in (Adger, 2003; Hornstein et al., 2005; Cook and Newson, 2007). The specifics of what exactly locality amounts to is of no importance to us, as this is not a text in linguistics. What's important is Chomsky's use of the concepts of 'rule', 'principle', 'parameter', 'universal', 'internal', 'innate', and so forth. These are the philosophically relevant concepts for our present purposes.

for instance, the subject can but need not be overtly expressed pronominally, as it is contained in the verb. So, the sentences

(7) Yo hablo Español
(8) Hablo Español
(*I speak Spanish*)

are both grammatically correct. In (7) the first person pronoun *yo* is overtly expressed whereas in (8) it is omitted without a change in the grammatical correctness and semantic import of the sentence since Spanish conjugates verbs according to *both* person (first, second, or third) and number (singular or plural) as in *hablo*, the first-person singular form of the verb *hablar*. The overt use of the pronoun *yo* in (7) is, from the standpoint of grammar, redundant[315] but permissible. Consider now what would happen if the two sentences were equally expressed in English,

(9) I speak English
(10) *Speak English

We know that (9) is a well formed, grammatically correct sentence whereas (10) is not. In fact, even if context settles who the subject of (10) might be, grammar doesn't. It is an ill-formed sentence in the grammar of English. We see that, evidently, in English the overt expression of the subject by means of the suitable pronoun is necessary or obligatory for the grammatical well-formedness of the sentence. The pronoun *I*, unlike its equivalent in Spanish *yo*, cannot be omitted in English if the sentence is to be grammatically well-formed. The parameter-setting model explains this difference by appeal of what is called the 'Null Subject Parameter', which in Spanish and English is set differently (cf. Cowie, 2008, p. 6). Thus, though both languages are constrained by the fact that sentences require a subject, in Spanish the subject parameter is set to 'Optional' whereas in English it is set to 'Obligatory'. Proposals as to how this setting of a parameter in one way as opposed to another takes place have been proposed and vary greatly. Fortunately, these are simply beyond the scope and point of our discussion (for more, see e.g. Ayoun, 2003; Pinker, 2015 ch. 3; Roeper and Williams, 1987). The simple, rough sketch outlined above should suffice to give an idea of how the principle and parameter (P&P) model is supposed to work. Earlier models (Chomsky, 1957, 1965) would have proposed different explanations, but they would have appealed to the innate constraints of the language faculty as contributing in some way to such differences as the one we just explored between Spanish and English.

UG principles are universal not because they are present *in the same respect or manner* in all languages, but rather because they are present in all languages in some way or another (which, in the P&P model, is a fact determined, partly, by the valuation given to prefixed parameters of variation) and are part of the innate structure of the language faculty of every human being. A language without question-movement will be a language without locality conditions for interrogative sentences, that is, without the principle of locality for questions. But this does not mean that it breaks or violates the principle, it simply means that the principle is absent in interrogatives in the relevant language because question movement does not take place in the relevant language. If, however, the language (say, Japanese) were to require wh-movement (or its equivalent), then if the locality hypothesis is true, it must involve locality conditions and therefore the movement must be

[315] Though not from the standpoint of pragmatics, or competent performance, where the pronoun *yo* can be used emphatically in a given context by someone to, say, bring home the point that she as opposed to others there speaks (proper) Spanish.

the shortest possible. A principle is UG universal if it is attributable to the language faculty rather than to some regularity or systematic pattern found in samples of data collected over many or all languages *contra* E-universals.[316]

Recall that for Chomsky, an E-language is a particular language in use or a grammar constructed by a linguistic on the basis of actually produced sentences. Examples of E-languages are Spanish, English, Dutch, and so forth, *as we encounter them in our everyday way of life* (e.g. in the physical manifestation through what is actually produced). But the actual use cannot account for grammatical competence,—knowledge of *what can and could be done* in the language and of what would be the case if what could be done were done. An I-language, on the other hand, as in the mastery of the grammar of English is a state of the language faculty that cannot be identified with the language in performance, though it is part of the conditions for the possibility for it, and it is the psychological result of a valuation and activation of a subset of the set of principles and parameters innately present in the language organ (the mind/brain) of the human person. Thus, every human has the same universal grammar (i.e. the very same principle and parameters structure) but not the same I-language (i.e. the activation of the same principles and the same valuation of the parameter variables).[317] Linguistics, Chomsky believes, should be concerned primarily with the I-language if it hopes to give a suitable answer to the question 'what constitutes knowledge of a language?'

Language is "a system represented in the mind/brain of a particular individual" (Chomsky, 1988, p. 36) rather than essentially a social phenomenon; therefore "the shift of focus from the dubious concept of E-language to the significant notion of I-language was a crucial step in early generative grammar" and, in particular, in the development of linguistics (Kasher and Chomsky, 1992, p. 10). Notice the recurrent ambiguous expression 'mind/brain' and the commitment to the 'internal/external' dichotomy in the analysis of the human mind. In this sense, Chomskyan linguistics faces the same problems that any internalist theory of the mind faces: bridging the gap between the internal and the external. But for present purposes our point of contention with Chomsky lies elsewhere: UG's theoretical commitment to the existence of a formal system in the mind/brain that is innate and is biologically determined in some yet unknown way.[318] The analogy is usually made in comparison with the human visual system, which is innate and has a definite

---

316 UG universals can be understood better by comparing with Greenbergian universals (named after "E-language" linguist Joseph Greenberg). The latter are postulated on the basis of regularities found in large samples of linguistic activity and it is arrived at inductively, whereas the former can be and have been postulated on the basis of analysis of a single language. The fact that there might be a language that does not contain a UG principle does not disprove that the principle is not a universal, since 'universal' in UG theory does not mean 'present in all languages' but rather 'present in the innate structure of the language faculty or mind'. Some languages, for a host of historical and developmental reasons concerning the given languages, do not use a given UG principle. But the absence of use is not proof of the inexistence of the principle. But for Greenbergian universals, so claims the UG theorist, the story is otherwise: one example of a language that does not use and adhere to the postulated Greenbergian universal is enough to refute the relevant Greenbergian theory. UG universals can be absent in several languages in one way or another in certain areas of a language, but they cannot be broken or violated. If they are, then this is empirical evidence that the postulated universal is not indeed a true universal principle of the innate structure UG. An example would be a natural language were movement in interrogatives takes place but not in the shortest way possible but rather, say, in the longest way possible.

317 Simplifying, we can say that two English speakers will have the same I-language but an English speaker and a Chinese speaker (assuming both are monolingual) will not have the same I-language though they will have the very same type of language faculty and therefore the same innate universal grammar structure. Chomsky's response as to why this is so is, in part, because we belong to the same biological species and therefore have all the same biological endowment (absent genetic anomalies and neurological impediments). The attentive reader should see how KRA poses a real problem for Chomsky's claim if the claim is that our unique neurobiology suffices to determine an innate formal structure, namely universal grammar.

318 For more advanced and technical introductions to Chomskyan generative grammars, see (Adger, 2003; Hornstein et al., 2005)

structure which develops as the child develops and adjusts to the suitable environment in which it is embedded. The structure of the visual system precedes any relevant modes and adjusted modes of visual perception and is that about *us* in virtue of which vision is possible. It is not a *tabula rasa*.[319] Analogously, Chomsky claims that the 'language organ' is neither a tabula rasa but a richly structured grammatical system of constraints (principles and parameters) called 'universal grammar' which makes the acquisition of any particular language possible.

Chomsky's postulation of a universal grammar in terms of tacit knowledge (or 'cognizance', as he later called it in order to avoid the "philosophically loaded" term 'knowledge'; see Chomsky, 1980, p. 92) of innate constraints (e.g. in the P&P model cashed out in terms of principles and parameters; in the MP model *inter alia*, in terms of computational mechanisms for recursion, etc.) in the language faculty located in the brain is his answer to the problem of which properties of us are the ones in virtue of which we can acquire and master a natural language and become grammatically competent with respect to a particular grammar.[320] Universal grammar is, psychologically speaking, a set of abilities not acquired for carrying out certain abstract computational operations that are grammatically relevant which comes built-in with our species specific neurobiological makeup; "universal grammar is part of the genotype specifying one aspect of the initial state of the human mind and brain." (Chomsky, 1980, p. 82)[321]

1.4 Chomsky's Conjunction and The Language Faculty

Our previous two chapters made the case for the claim that formal understanding, the ability to carry out formal operations cannot be a wholly physical (e.g. biological) phenomenon. Indeed, we showed that the capacity for that sort of ability and activity could not be wholly material (chapters 2 and 3). We've also made it clear that this does not imply that our formal capacities, abilities, and activities do not require, as a necessary condition, a suitable material basis (e.g. our species-specific biological endowment, material means of representation, etc.). We have merely argued against its sufficiency in regard to both the genetic and constitutive account of our formal capacity and our

---

[319] The analogy is misleading. One need not deny that there is a biologically determined structure to the visual system to claim that the visual system is a tabula rasa with respect to *vision*. A realist can claim that though human beings have been endowed with a visual apparatus suitably structured for the relevant environment it does not contain any of the features of what is *seen*—let alone innately and inborn. Thus, he would claim that it makes no sense to say that there is some universal vision in the visual system. One could make a similar argument against Chomsky, but it is not one we wish to make here.

[320] It is, in fact, an answer to two interconnected problems: the problem of the poverty of stimulus and the problem of creativity. Roughly, the former concerns the fact that a child is able to comprehend the principles and rules of a language that apply to infinitely many sentences of the language on the basis of an extremely limited, impoverished finite sample of linguistic activity or data. At a point in her development, the child displays the possession of an ability that extends way beyond anything that could be given by the primary linguistic data (*pld*), which is the immediate, impoverished, and finite linguistic activity and behavior to which the child is exposed while growing up. Chomsky's claim is that the *pld* (1) do not contain all the grammatical rules and principles which children display while mastering a natural language (say, their mother tongue) and therefore (2) cannot suffice to account for the linguistic ability of children. He then concludes that the most reasonable hypothesis is that the ability precedes the *pld*, and that there is an innate universal grammar in the mind/brain that constitutes the child's ability to know any particular natural language. The latter we have already explained: it is the ability to comprehend (at least grammatically) and generate ever new and never before encountered sentences. This too, claims Chomsky, cannot be explained on the basis of *pld* and therefore must precede it and be constitutive of the structure of the mind.

[321] Because Chomsky takes language to be part of the genetic inheritance and makeup of human beings, even perhaps with a particular organ (the language organ), there exists (at least epistemically) the possibility of individual variation in language much like there is variation in the color of the eyes and skin, in the color and kind of hair, height, bone structure, and so on, amongst individual members of the same species.

formal ability and activity. But our use of 'formal understanding' with respect to language has been somewhat ambiguous in the previous chapters. This is because it is not immediately clear whether linguistic understanding is a species of formal understanding or whether it is the condition for the possibility of and constitutive of formal understanding. I believe the latter option is closer to the truth of the matter whereas the former would seem to be closer to Chomsky's philosophical approach to the nature of linguistic understanding. But that is not to be settled here. As we've already explained, there are some linguists who posit pure formal structures (whether in terms of rules, or functions, or principles and parameters) as innate conditions for the possibility of acquiring a particular language and as constituting the formal-structural underpinning of any individual's knowledge of a particular language (what these linguists call an 'E-language' as opposed to the 'I-language'). The problem I wish to bring to the fore does not have to do with nativism about language (i.e. that language is in some sense innate) but rather with the conjunction of nativism and neurobiologism about the language faculty if the language faculty is, at least in part, formally constituted.

Chomsky is known and credited for having done away with behavioristic approaches to the study of language and having propelled the cognitivist turn or revolution in linguistics. He is known for having proposed the theory of UG which posits a grammar (a set of rules or a set of principles and parameters, depending on the model) that is innate and inborn to every human being, which every human being tacitly knows and unconsciously employs, and which grounds the range of possible existent languages as well as the acquisition of any particular grammar or language. The structure of UG is abstract, but it is supposed to be manifested in particular grammars like English, Spanish, Dutch, and so forth, as *internalized*. As of this date there have been at least six different UG models proposed by Chomsky, nearly one per decade and all with intervening phases, for the study and modeling of the language faculty and its innate structure. [322] It's impossible to canvass the history of generative grammars here, but the problem we wish to raise should be clear enough: if the faculty of language is (even partially) computational or, more generally, a formally definite structure then neither the capacity for language nor mastery of a language can be given a constitutive and genetic explanation that is wholly neurobiological. Insofar as UG theory posits abstract, formal structures as constitutive of the language faculty, that faculty cannot be wholly material and therefore the brain (read: those core parts of the brain that enable and are involved *as necessary conditions* in language acquisition and competence) cannot be sufficient to constitute the formal structure of the language faculty. In short, Chomsky's conjunction cannot be consistently held. To this we now turn.

---

[322] They can be broken down by decades. The original model became known as Syntactic Structure, which took its name from the title of (Chomsky, 1957). This model was then followed and superseded by a model proposed in (Chomsky, 1965), which became known as the Aspects Model or the Standard Theory. This model, the Standard Theory, then morphed into the Extended Standard Theory (EST) during the 1970s, which later was radically changed into the Government/Binding (BD) Model which began with (Chomsky, 1981, cf. 2010) and was later developed in (Chomsky, 1986). The notion of principle and parameters was introduced with BD and given its foundational and theoretical prominence over other notions of BD is has come to be known as the Principles and Parameter (P&P) Theory. Then there is the Minimalist Model (Chomsky, 1993b, 1995b, 2000) which has had about three phases thus far. Needless to say, each model generated a proliferation of different theoretical approaches, refinements, formal machinery, and theories within a given model. It is impossible to consider in depth not only each model but even a single model of transformational generative grammars here. It would require a further dissertation. Our aim here is quite minimalist (no pun intended), as outlined in at the beginning of this chapter, for we focus on what is philosophically perennial and continuous to Chomsky's approach to linguistics rather than the array of differences between (and within work carried out) in all the models or frameworks he has proposed.

2. KRA AND CHOMSKY'S CONJUNCTION

In terms of capacities, we can state the implications of KRA as follows,

1. Formal capacities (i.e. the capacity for the ability to follow and apply formal rules or compute pure functions) *cannot* be wholly physical capacities. (Contra type-identity theories)
2. Formal capacities *cannot* be the (actual or potential) physical realization or implementation of formal structures. (Contra token-physicalist theories whether functionalist, computationalist, connectionist, etc.)
3. Formal capacities *cannot* be wholly determined by nor can they be wholly the outcome of anything physical. (Contra emergentist and evolutionary theories)
4. Formal capacities *cannot* be wholly contained nor stored in anything physical. (Contra materialist storage hypotheses, say, formal rules are literally stored by/in the brain)
5. Formal capacities are real and inherent in whatever has them. (Contra eliminativists and instrumentalists)[323]

Filling in some missing premises and *mutatis mutandis*, we can get the same implications for 'formal abilities' and 'formal activities'. The qualification 'wholly' is important, because there are surely biologically determined (i.e. neurological, sensory) processes that act as part of the material mechanisms underlying the relevant abilities and their exercise (i.e. the activity). Likewise, there would be no formal understanding if there were no memory (e.g. formal understanding can be lost due to injury or forgotten simply due to lack of engagement with the relevant concepts) and memory is grounded in and a feature of our neurobiological nature. Again, a human person is a psycho-somatic unity, and a proper account of that unity that leaves room for the non-physicality of formal understanding without sacrificing embodiment as a necessary condition for it must be developed and explored if KRA is sound. I take this (i.e. the material conditions for human cognition) to be a *sine qua non* condition for any explanation of the intellectual capacities and abilities of human beings, though substance dualists may beg to differ. This must all await another occasion. For now, we can be sure that a materialist account, whether reductive or non-reductive, of any formal capacity, ability, and activity is untenable. It would make the formal aspects of the human psyche unintelligible and any formal ability and activity an impossibility. Moreover, it is self-refuting (see chapter 5).

The language faculty is not merely formally determinable but also formally determinate in the ways in which the relevant UG theory says it is. Language is something that 'grows' out of the universal grammar of the brain rather than something that is acquired. The generative grammarian could object that Kripkestein's considerations as well as our argument are anachronistic and miss their mark insofar as they depend on the notion of 'rules', which the UG theoretical framework has long since abandoned as fundamental. As of the 1980s and in particular during the 1990s, Chomsky

---

[323] Note that we have not agued at length or explicitly for either 4 or 3, respectively. The issue of causation, and therefore emergence, was touched upon briefly in chapter 5. But other than that, we have left it to the reader to see how the considerations that give way to KRA are easily applicable to causal and emergentist theories of formal capacities and abilities insofar as they rely on a suitable structured material basis as the sufficient ground for the coming about or the emergence of formal capacities or formal abilities. Our argument explicitly addressed and argued for propositions 2 (chapters 2 and 3) and 5 (chapters 4 and 5, *contra* Dennett). Proposition 1 is readily refuted by the considerations given in chapters 2 and 3 which are, in fact, easier to apply to any potential type-identity theory of formal operations. Proposition 4 we take to follow from propositions 1, 2, 3 and 5. Thus, we take our argument to have established if not all the five propositions, at least the ones that lay the groundwork for showing thoroughly why the other propositions follow or hold. We leave this for another occasion.

moved away from the conception of language knowledge as rule-based and linguistic performance as rule-governed activity to a more abstract, general characterization of language knowledge that minimizes the role of rules and subsumes them under a more fundamental category of language determinability: principles and parameters (P&P), from which individual rules are said to be derived. "The basic assumption of the P&P model of language is that languages have no rules at all in anything like the traditional sense, and no grammatical constructions (relative clauses, passives, etc.) except as taxonomic artifacts." (Chomsky 1995b, p. 388)[324] Rules still play a role at the level of particular grammars, but they must be accounted for by a set of principles and parameters at the level of universal grammar that accounts for their existence (say, as secondary to and derivative from principles and parameters). As Cook and Newson explain,

> The information stated in rules has to be reinterpreted as general principles that affect all rules rather than as a property of individual rules. Rules are by-products of the interaction between the principles and the lexicon [one of the vital components of the computational system that constitutes language knowledge and consists of lexical entries that contain a mass of information about how the lexical item, i.e. the word, in question behaves grammatically]. UG Theory is not concerned with specific syntactic constructions such as a 'passive' or 'relative cause' or 'question,' or the rules which linguists can formulate to express regularities in them, which are simply convenient labels interactions of principles and parameters. The passive is not an independent construction so much as the product of a complex interaction of many principles and parameter settings, each of which also has effects elsewhere in the syntax: "a language is not, then a system of rules, but a set of specifications for parameters in an invariant system of principles of Universal Grammar (UG)." (Chomsky, 1995b, p. 388)
>
> (Cook and Newson, 2007, p. 9)

But this is not to the point. KRA, like the skeptic's argument, can be formulated in terms of the formal principles and parameters of UG theories. For any principle Σ and parameter Ξ that some UG theory posits as constitutive of the universal grammar in the brain, there will at least some other incompossible principle Σ' and parameter Ξ' that some qu-UG theory posits that the neurobiological and genetic makeup of the human being will inevitably satisfy. What can be said of rules using our argument can be said of any formal structure that requires formal operations, such as P&P.

Consider the now outdated and superseded rewrite rules discussed in chapter 1.

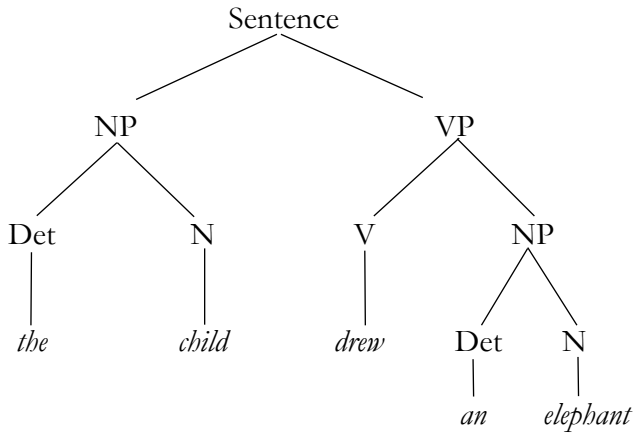
(1) The child drew an elephant.

There are two phrases in this sentence: the noun phrase (NP) *The child* and the verb phrase (VP) *drew an elephant*. Like truth trees, the tree diagram of a sentence breaks its phrase structure down to its lowest, atomic level. In the case of (1), for example, the VP breaks down into a verb (V) *drew* and an NP *an elephant*, whereas the NPs in turn break down into a determiner (D or Det), *the* and *an*, and a noun (N), *child* and *elephant*, respectively. The tree diagram of (1) then looks like this:

---

[324] "There has been a gradual shift of focus from the study of rule systems, which have increasingly been regarded as impoverished, … to the study of systems of principles, which appear to occupy a much more central position in determining the character and variety of possible human languages" (Chomsky, 1982, pp. 7–8)

(S)



The actual words at the bottom can be substituted for indefinitely many other words to form a new sentence with the same structure, i.e. phrase structure, so long as the word-substitutions are of the right type.[325] The phrase structure of (1), call it (S), is repeatable and instantiable by infinitely many possible sentences. (S) can just as well be represented using labelled bracket, for example

(S´) [$_S$ [$_{NP}$ [$_D$ *The*] [$_N$ *child*]][$_{VP}$ [$_V$ *drew*][$_{NP}$ [$_D$ *an*] [$_N$ *elephant*]]]].

This formal structure is what the person with linguistic understanding of English knows, being that on account of which she is able to generate ever new sentences that are in accord with the given abstract grammatical structure. There are also sentences that do not possess this structure, of course, but the point of illustrating this is to show that phrase structures are the abstract invariable component of infinitely many potential sentences.[326] What Chomsky did was formulate a simpler and more powerful form of representation of phrase structures, called rewrite rules. So (S) is a sentence that follows the following syntactic rules:

(Si) S → VP NP
(Sii) VP → V NP
(Siii) NP → Det N
(Siv) NP → N

where the symbol '→' stands for 'consists of', 'S' for 'sentence', and so forth. One then goes on to provide the same rewrite rule to all constituent expressions (e.g. VP and NP).[327]

---

[325] I abstract away from the constraints placed on the possible instances of a type of syntactic structure by the number of nouns, verbs, and determiners in the lexicon of the English language. For our purposes, nothing of importance hinges on this.

[326] To be sure, the notion of phrase structure in linguistics goes back to the 1930s and precedes the advent of Chomsky's transformational generative grammars. I am here interpreting them in light of the 'cognitivist turn' in the late 50's due to (Chomsky, 1957, 1959). "This proposed format [phrase structure rules that form phrase-markers together with transformational rules that convert phrase markers into other phrase-markers] was adapted from traditional descriptive and historical grammar, recast in terms of ideas developed in the theory of computation (recursive function theory, theory of algorithms)." (Chomsky, 1986, p. 56)

[327] We ignore lexical rules and rules used to associate phrase structure representations to phonetic form. It is not to our point, though we could make our case with respect to these just as well if these are taken to be computational or algorithmically determined.

As we've already said, our exposition is a crass simplification of the more complex and technical formal machinery introduced in (Chomsky, 1957) and thereafter for the rules for generating (and transforming) infinitely many well-formed grammatical sentences of a certain syntactic type. Theoretical generative linguistics has in general moved way beyond the characteristics of this early work on transformational generative rules, particularly the transformational aspect and the abandonment of the concepts of rule, theory, and hypothesis in the specification of the innate universal grammar (for the culmination of the shift from the early transformational generative rule-governed approach to language to a principle and parameter generative one, see e.g. Chomsky, 2010, 1981, 1980; though talk of "rules", in the idiosyncratic way in which Chomsky understands that concept, was not entirely abandoned).[328] I have chosen phrase structure rules for matters of simplicity and because it suffices for our present purposes. What is of importance is to note how a potential infinity of well-formed grammatical sentences can be generated from a finite set of rules (that is, a finite set of rules for constructing grammatically well-formed syntactic structures) and thus how the creativity aspect of language knowledge, however imperfectly, is captured even by the phrase-structure format (e.g., one knows that any sentences constructed out of those rules will be a grammatrically well-formed sentence, a sentence of a correct syntactic type, regardless of the specific variable components, which appear at the bottom of the tree diagram, so long as they are of the right sort). It also neatly enables us to express a fundamental feature of Chomsky's project that has survived all his proposed modifications to the framework of transformational generative grammars: that there are *inter alia* computational procedures, a set of definite algorithms, for generating types of syntactic structures independently of any specific variable component (again, so long as the variable component is of the right *type*).

---

[328] Chomsky's original rewrite rules were found to be language-specific constructions and therefore lacking in the sort of generality and universality required by the descriptive and explanatory standards of UG theory (more abstract and general strategic attempts, like the introduction of X-bar theory (Chomsky, 1970b; Jackendoff, 1977) to mention just one example, were made in order to mitigate the implosion in complexity of phrase structure systems on account of the immense variety of language-specific rules and constructions; the strategy beginning around the 1960s was to take the immense variety and diversity of rules and construction to be superficial that could be more basic general principles could be found to which all rules adhered to). One the of the fundamental theoretical goals, however, was and is the same: to account for language's power to generate a potentially infinity of discreet grammatical units (e.g. syntactic structures of a certain type, iterations of sub-clauses, etc.). So, to quote the man himself, in the early days of generative grammar there had been "an attempt to come to terms with a kind of a paradox that emerged as soon as the first efforts were made to study the structure of language very seriously, with more or less rigorous rules, an effort to give a precise account for the infinite range of structures of language. The paradox was that in order to give an accurate descriptive account it seemed necessary to have a huge proliferation of rule systems of a great variety, different rules for different grammatical constructions. For instance, relative clauses look different from interrogative clauses and the VP in Hungarian is different from the NP and they are all different from English; so the system exploded in complexity. On the other hand, at the same time, for the first time really, an effort was made to deal with what has later come to be called the logical problem of language acquisition. Plainly, children acquiring this knowledge do not have that much data. In fact … it's very limited; still, somehow children are reaching these states of knowledge which have apparently great complexity, and differentiation and diversity – and that can't be. Each child is capable of acquiring any such state; children are not specially designed for one or the other, so it must be that the basic structure of language is essentially uniform and is coming from inside, not from outside. But in that case it appears to be inconsistent with the observed diversity and proliferation, so there is kind of a contradiction, or at least a tension, a strong tension between the effort to give a descriptively adequate account and to account for the acquisition of the system, what has been called explanatory adequacy." (Chomsky, 2002, pp. 92–93) Let me point out I have intentionally omitted any mention of the early *transformational* rules (e.g. rules for transforming one type of syntactic structure, like sentences in the active voice, into another type of syntactic structure, the passive) that were part and parcel of the early work on generative grammar (originally called 'transformational generative grammar'). The notion of 'transformational' has figured less and less in the work of Chomsky through the years, most noticeably *post* 1980.

Since the 1980's Chomsky has gradually moved away from talk of language and the language faculty as rule-governed and an innate system of generative and transformational rules to specifications and valuations of principles and parameters[329] which are universally shared by language speakers, the difference between languages lying not so much in the set of principles but in the choice between, or the valuation given to, the innately fixed variables or options, called *parameters*.[330] (Think of the language learner as a switchbox (Chomsky, 1988, pp. 61–62) whose switches are all in place, i.e. come built-in, but can be switched on or off or flicked to one position or another. Then think of certain switches going on and others remaining off in the switchbox that is the universal grammar in the mind/brain of the child as she is exposed to a pattern of 'switching' or 'flicking' the innate parameters to one position or another, the result being a parameter-setting occasioned by the linguistic data she is presented with during the course of her growth, development, and maturation in a given linguistic environment.) But the general *philosophical* principle remains operative: that the nature of language (or the nature of the capacity for language) is, at least in part, formal and therefore is constituted partly by a pure formal invariable component (universally shared by all human beings by natural endowment) that is the subject-matter of formal (e.g., generative) approaches to language and whose actualization lies in the application of highly abstract, formal (generative) principles, parameters, and operations (e.g. computational mechanisms for recursion). If this is so, then our argument will have to apply to the capacity for language (i.e., language faculty) and to what is attained in virtue of the interplay of that capacity with a host of

---

329 "What happened at Pisa [later published as (Chomsky, 1981)] is that … a method arose for sort of cutting the Gordian knot completely: namely eliminate rules and eliminate constructions altogether. So you don't have complex rules for complex constructions because there aren't any rules and there aren't any constructions. There is no such thing as the VP in Japanese or the relative clause in Hungarian. Rather, there are just extremely general principles like "move anything anywhere" under fixed conditions that were proposed, and then there are options that have to be fixed, parametric choices: so the head of the construction first or last, null subject or not a null subject, and so on. Within this framework of fixed principles and options to be selected, the rules and the constructions disappear, they become artifacts." (Chomsky, 2002, p. 94)

330 Take movement, unqualifiedly. Whether movement so taken is a universal principle is an empirical hypothesis (e.g. do all languages use movement in at least one respect?). Suppose movemet is a universal. We know that it will not manifest itself in the same respect in every language. So, for example, not all languages have question or wh-movement, though they must all have movement of some sort if movement *as such* is a genuine universal. This is a parameter of variation. All languages have movement but not in the same manner, due to the triggering or use of differing parameters. What cannot be a parameter of variation is that if movement is present, in whichever respect, it must be the shortest possible *if* the locality principle of movement is truly an UG principle. The hypothesis that movement should be the shortest possible is about certain conditions and properties that all movements must have. The fact that movement in questions is present in English but not in Baha Malaysia and Japanese questions does not disprove that movement is a universal principle but merely that movement is affected by parameters of variation. Otherwise all languages would the same. The principle may absent in some languages while present in others with respect to certain sorts of grammatical structures or constructions. UG tells us that to not make use of a principle is not to break or violate the principle. One thing is to have a language that does not require anything like wh-movement in interrogatives, another is to have a language that does require it and makes it the longest possible (this would be a violation not of the principle of movement but of the principle or constraint of locality of movement). Note that there is a "gap" between the language faculty and UG theories: UG theories are *theories about* the innate structure of the language faculty whereas the innate structure of the language faculty is what every member of the human species has independently of any particular UG theory. UG as a theoretical program, however, posits and is committed to the real existence of such innate structure and so, though it might get things wrong when it comes to the precise structure of the language faculty (say, a UG theory in the vein of the P&P model that formulates a principle that turns out not to be a principle of the language faculty but a regularity of a specific language), it does not get it wrong when it comes to the existence of an innate structure in the language faculty. The task of UG theories is to discover, or uncover, the innate structure of the language faculty and explain how it gives rise to (generates) particular grammars. Thus, the expression 'universal grammar' can at times be used to refer to the innate structure of the language faculty and at other times to the linguistic theories and models of such innate structure. For an acknowledgment of this ambiguity in the literature, see e.g. (Chomsky, 1986, pp. 28–29)

internal-external factors (e.g., maturation, unconscious operations, the linguistic data and input the child receives from members of the linguistic community, environmental factors, etc.). That is to say, linguistic understanding and linguistic activity. If this is so, Chomsky's neurobiologism about the language faculty stands refuted. So, Chomsky faces a dilemma: either he abandons a commitment to the formal nature of the language faculty, namely that the language faculty is formally rich and determin*ate* [331] (and, therefore, renounces the idea that there is a universal *grammar* in the language faculty) or he abandons his neurobiologism (in which case he can retain the formal structure of the language faculty but must abandon his claims that the innate structure of the language faculty is wholly neurobiologically determined and that UG theories are on a par with the theories in chemistry or biology).

Again, lest I be misunderstood, I am not arguing against nativism about language, even if what is native or innate is taken to be a determinate formal structure of some sort, nor am I arguing against Chomskyan linguistics as such. I am, instead, arguing against a crucial aspect of the *philosophical* framework underlying Chomsky's approach to language and linguistics, namely what I have called Chomsky's conjunction. I am interested therefore in whether the formal structure, be it innate or otherwise, which Chomsky calls universal grammar can be (1) identified with a material structure (e.g. the human brain and nervous system), (2) identified with a structure that is materially realizable in an intrinsic way on account of some material properties and configuration, (3) identified with emergent and non-reducible structures that are nevertheless wholly determined and caused materially, or (4) contained in some material locus like the brain. KRA shows that all of these proposals all fail to meet the intelligibility conditions (e.g. that the application of a formal rule can be formally definite or determinate) for being a suitable subject of definite formal properties and features. All of them violate (DR) and the principle of exclusion. This is a problem for Chomsky insofar as he is committed to one or more of the proposals from (1) to (4). So we find claims in the literature all in the spirit of Chomsky to the effect that UG "is a theory of linguistic invariance, as it expresses the universal properties of natural languages…" that "expresses the biologically necessary universals, the properties that are universal because they are determined by our in-born language faculty, a component of the biological endowment of the species." (Belletti et al., 2002, p. 9)

Note that since the 1990s, as part of the Minimalist Program (MP), Chomsky (Chomsky, 1993b, 1995b, 1995a, 2000, 2001) has devoted particular attention to the purported interface (that is, to bridging the gap) between the 'outside' world of sounds and the 'inside' world of concepts, both of which are not language. Language knowledge is identified with a computational system, namely the computational system that links the outside world by processing and converting according to certain rules the external (e.g. phonetic sounds) into the internal (e.g. meanings, concepts) and the internal into the external. There's also the development that recursion is *the* significant trait of the language faculty. The details of the theory are of no significant importance for our present purposes. What matters is the metaphysical and ontological categorization assigned by Chomsky to the structural components of the language faculty. KRA makes it clear that if language knowledge is a

---

[331] Actually, this is misleading as one can hold that natural languages have a formal nature or structure without also holding to UG as the proper model or theory for explaining that structure or even to the claim that the capacity for language is innately formally rich and determinate. So, Chomsky could renounce UG while still maintaining a commitment to the formal nature of language and linguistic understanding. But even if Chomsky abandons UG, this does not get him off the hook unless he abandons either a commitment to the claim that languages have a formal structure or a commitment to neurobiologism about language (or both). Of course, if Chomsky renounces instead his commitment to neurobiologism then he can hold to the formal nature of language and even UG. But he cannot have both a commitment to the formal structure of language (the language faculty, that is) and a commitment to neurobiologism about language (the language faculty). The commitment is incoherent and self-refuting (see chapters 4 and 5).

computational system (e.g. computational mechanisms for computing recursive functions of the relevant sort), then the relevant formal rules and functions cannot be implemented nor determined in their entirety by material processes, including any species-specific neurobiology. Matter alone cannot suffice for formal determination.[332] So this is our claim: if the nature of universal grammar is, at least partly, formal then it cannot be wholly explained by appeal to any material basis and cause, be it biological or otherwise.

3. CHOMSKY ON RULES AND RULE-FOLLOWING

Interestingly, Chomsky (Chomsky, 1986, pp. 221–275) addressed Kripke's skeptical paradox and the rule-following considerations. Kripke was sensitive to the fact that Wittgenstein's rule-following considerations were relevant to and consequential for Chomsky's generative approach to the Central Project of theoretical linguistics. There are three main instances in which Chomsky's UG theory comes up in Kripke's discussion, and they come up in footnotes (rather long footnotes). Before delving into Chomsky's critique, it will be useful to quote one of the passages in which Kripke discusses some of the repercussions that Wittgenstein's rule-following considerations may have for Chomsky's approach to the nature of language and linguistic understanding. We will follow this up by presenting Chomsky's critique of Kripkestein. But I am not here going to defend Kripkestein from Chomsky; rather, I shall consider Chomsky's response as a possible response to KRA and show that it fails. Therefore, though Kripke does not seem to have further developed elsewhere his remarks on Chomskyan linguistics (not even in his unpublished work on computationalism), we will explore how the argument can be developed so as to apply to UG theories together with how KRA affects the philosophical presuppositions of Chomsky's Cartesian linguistics.[333]

Kripke points out that Chomsky's concept of competence is consequent upon and unintelligible without the concept of rule-following and is therefore subject to the very same Wittgensteinian considerations about rule-following.

> ... what is important here is that the notion of 'competence' [or 'knowledge of a language', in Chomsky's jargon] is itself not a dispositional notion. It is normative, not descriptive, in the sense explained in the text. The point is that our understanding of the notion of 'competence' is dependent on our understanding of the idea of 'following a rule', as is argued in the discussion above. Wittgenstein would reject the idea that 'competence' can be defined in terms of an idealized dispositional or mechanical model, and used without circularity to explicate the notion of following a rule. Only after the sceptical problem about rules has been resolved can we then define 'competence' in terms of rule-following. Although notions of 'competence' and 'performance' differ (at least) from writer to writer, I see no reason why linguists need assume that 'competence' is defined prior to rule-following. Although the remarks in the text warn against the use of the 'competence' notion as a solution to our problem, in no way are they arguments against the notion itself. Nevertheless, given the sceptical nature of Wittgenstein's

[332] "[…] whether language is innate or not is, after all, an empirical issue." (Cowie, 2008, cf. 1999, 1997) Is it? One can also raise the point that it is an empirical issue only if it *makes sense* to say that linguistic or grammatical rules and structures can be innate, and in particular innately present in the brain (that is, in a way that is empirically testable or reachable). This is an open question.

[333] The label 'Cartesian Linguistics' is Chomsky's (Chomsky, 2009), the title of a book originally published in 1965. He broadly identifies his project in linguistics with the long-standing Cartesian rationalist tradition in epistemology, philosophy of mind, and metaphysics.

solution to his problem … it is clear that if Wittgenstein's standpoint is accepted, the notion of 'competence' will be seen in a light radically different from the way it implicitly is seen in much of the literature of linguistics. For if statements attributing rule-following are neither to be regarded as stating facts, nor to be thought of as explaining our behavior … it would seem that the use of the ideas of rules and of competence in linguistics needs serious reconsideration, even if these notions are not rendered 'meaningless'. (Depending on one's standpoint, one might view the tension revealed here between modern linguistics and Wittgenstein's sceptical critique as casting doubt on the linguistics, or on Wittgenstein's sceptical critique or both.) These questions would arise even if, as throughout the present text, we deal with rules, like addition, that are stated explicitly. These rules we think of ourselves as grasping consciously; in the absence of Wittgenstein's sceptical arguments, we would see no problem in the assumption that each particular answer we produce is justified by our 'grasp' of the rules. The problems are compounded if, as in linguistics, the rules are thought 'of as tacit, to be reconstructed by the scientist and inferred as an explanation of behavior. The matter deserves an extended discussion elsewhere. (See also pp. 97 to 99 and n. 77 below.)

(Kripke, 1982, pp. 30–31 n. 22)[334]

---

[334] The other two passages are:

> Donald Davidson's influential and important theory of natural language has many features in common with the *Tractatus*, even if the underlying philosophy is different. Davidson argues that some simple, almost a priori considerations (not requiring detailed empirical investigation of specific natural languages) put strong constraints on the form of a theory of meaning for natural languages (it must be a finitely axiomatized Tarski-style theory of truth conditions). (Although the form of a theory is determined without detailed empirical investigation, for a particular language the specific theory adopted is supposed to require detailed empirical support.) The fact that a theory of meaning must have this form, it is argued, puts strong constraints on the logical form, or deep structure, of natural language - very probably that it ought to be close to classical extensional first order logic. All these ideas are close to the spirit of the *Tractatus*. In particular, like the *Tractatus*, Davidson holds (i) that truth conditions are a key element in a theory of language; (ii) that the uncovering of a hidden deep structure of language is crucial to a proper theory of interpretation; (iii) that the form of the deep structure is constrained in advance by theoretical, quasi-logical considerations; (iv) that, in particular, the constraints show that the deep structure has a logical form close to that of a formal language of symbolic logic; (v) that, in particular, sentences are built up from 'atoms' by logical operators; (vi) that, in particular, the deep structure of natural language is extensional in spite of the misleading appearances of surface structure. All these ideas of the *Tractatus* are repudiated in the *Investigations*, which is hostile to any attempt to analyze language by uncovering a hidden deep structure. In this last respect, modern transformational linguistics, since Noam Chomsky, has been closer to the *Tractatus* than to the *Investigations*. (But for transformational grammarians, even the form of the theory is established by specific empirical considerations requiring detailed investigation of specific natural languages.) See also the programs of the linguists who called themselves 'generative semanticists' and of Richard Montague. Of course many of the ideas of the *Tractatus*, or of 'logical atomism', have not been revived in any of these theories.
>
> (Note: In recent transformational linguistics, 'deep structure' has a specific technical meaning. 'Generative semanticists' made the repudiation of 'deep structure' a key plank of their platform. In the preceding, it is best to take 'deep structure' in the general sense of 'underlying' structure. Anyone whose theory of language leads him to applaud the doctrine of *Tractatus* 4.002 - that the understanding of language involves countless tacit conventions, invisible to the naked eye, that disguise form - believes in deep structure in this broad sense. 'Deep structure' in the specific sense was a special theory of deep structure thus broadly defined; that is one reason why it was an appropriate term. Most recent linguistic theories that rejected 'deep structure' in the specific sense accepted it in the broader sense.)
>
> (Kripke, 1982, pp. 71–72 n. 60)

Now recall that according to Chomsky, competence (mastery of the grammar of a language) is an internal state and property of the individual brain/mind (Chomsky, 1986, p. 224). But if the concept of linguistic/grammatical competence is consequent upon the concept of rule-following and mastery of grammatical rules, then the skeptic's argument can be made in a way that subjects Chomsky's view of language to the very same indeterminacy considerations: there is not fact of the matter as to whether someone is grammatically competent or has grammatical competence; for there is no such thing as an objective non-relative fact or set of properties of the individual person that determines whether she follows a definite rule or a definite set of rules *R* as opposed to some other incompossible rule or set of rules *R´*. The skeptic's argument implies that there is not fact of the matter about a person's psychology that determines a definite linguistic competence and so no fact of the matter as to whether anyone has grammatical competence. The proposed solution to the skeptical paradox (Kripke, 1982, p. 55ff) concedes precisely this much but argues for conditions under which a given community would take it as appropriate to assert that someone is following a rule *R* (that is, no truth-conditions, only assertability-conditions).

Chomsky's criticism of Kripkestein seldom figures in philosophical literature on rule-following. I suspect this is due to how defective Chomsky's attempted refutation of Kripke's argument is. At moments, Chomsky seems to display adequate understanding of the problems raised by Kripke's skeptic, sometimes expressing them quite eloquently, only to follow this feat with a completely irrelevant counter-argument. Here's Chomsky's summary of Kripke's skeptical paradox,

> Wittgenstein's skeptical paradox, in brief, is this. Given a rule *R*, there is no fact about my past experience (including my conscious mental states) that justifies my belief that the next application of *R* does or does not conform to my intentions. There is, Wittgenstein argues, no fact about me that tells me whether I am following *R* or *R´*, which coincides with *R* in past cases but not in future ones. Specifically, there is no way for me to know whether I am following the rule of addition or another rule (involving

---

> Modern transformational linguistics, inasmuch as it explains all my specific utterances by my 'grasp' of syntactic and semantic rules generating infinitely many sentences with their interpretation, seems to give an explanation of the type Wittgenstein would not permit. For the explanation is not in terms of my actual 'performance' as a finite (and fallible) device. It is not a purely causal (neurophysiological) explanation in the sense explained in the text; see note 22 above. On the other hand, some aspects of Chomsky's views are very congenial to Wittgenstein's conception. In particular, according to Chomsky, highly species specific constraints – a 'form of life' – lead a child to project, on the basis of exposure to a limited corpus of sentences, a variety of new sentences for new situations. There is no *a priori* inevitability in the child's going on in the way he does, other than that this is what the species does. As was already said in note 22, the matter deserves a more extended discussion.
>
> (Kripke, 1982, pp. 97–98 n. 77)

Though it may seem as if these passages point to the exact opposite of what I'm saying, this is only apparently so. Consider our main argument, KRA. Suppose the acquisition and the mastery of a language involved amongst other things the capacity and ability (even if unconsciously so) to carry out a set of formal (e.g. computational) operations, however simple or complex. KRA will only imply that Chomsky's neurobiologism about the capacity and ability to carry out the linguistically relevant formal operations is misplaced. But our argument remains silent with respect to many other aspects of Chomsky's approach to language (e.g. nativism, species-specific constraints on possible languages, computational mechanisms, etc.). Of course language is species-specific and of course there are species-specific constraints! Of course the capacity for language is inborn, and perhaps it is even a computational mechanism! Who knows? Our argument only refutes Chomsky's conjunction. It doesn't show which conjunct is false (not even whether both are) but rather that at least one conjunct has to be false and that therefore the conjunction as such cannot be true (nor coherent).

> "quus," not "plus") which gives the answer 5 for all pairs beyond the numbers for which I have previously given sums; "there was no *fact* about me that constituted my having meant plus rather than quus," and more generally, "there can be no such thing as meaning anything by any word." Each application of a rule is "a leap in the dark." My application of a rule "is an unjustified stab in the dark. I apply the rule *blindly*." The argument is not limited to use of concepts but extends to any kind of rule application.
>
> (Chomsky, 1986, p. 225)

Ignoring the fact that the skeptic's argument ultimately makes a *presentist* case (that no present or occurrent fact of the matter exists that determines whether one is applying R or R´), this is as far as it goes on the mark. In the last chapter, we saw how the considerations that lead to the skeptical paradox turn out to be incoherent and unintelligible when applied to human beings but wholly coherent and intelligible when applied to wholly physical, material phenomena. Now, Chomsky gets into shaky waters when considering the so-called solution to the skeptical paradox (Kripke, 1982, pp. 55–113), which he takes to be the most important challenge to his account of language. Consider the following propositions:

(1) There is no fact of the matter *about someone* that can determine whether she is following a formal rule (that is, whether she is following *R* and not *R´*).[335]
(2) There is no fact of the matter *about a community* that can determine whether someone is following a formal rule (that is, whether she is following *R* and not *R´*).

Kripkestein's solution to the skeptical paradox consists in the concession of (1) but the rejection of (2).[336] There is nothing intrinsic to the individual that determines whether she is following one rule rather than another but there is a community, with its established rule-following practices and 'forms of life,'[337] that provides conditions for asserting whether the individual person is following some rule *R* or some other rule *R´*. Naturally, different communities with rule-govened activities will have different assertability conditions, and one can be said to follow a rule *R* relative to a community *A* if one manifests the patterns of behavior that have been taken or 'codified', so to speak, by *A* as instances of applying *R*. Kripkestein's solution is to reject (2), replace 'truth-conditions' with 'assertability conditions', and say that which rule a person follows is determined by whether her application of the rule is sufficiently similar to how a given community would apply the rule (or sufficiently like the applications that the members of a community are inclined to). Communities with incompossible standards will attribute to the same person at a given time the application of incompossible rules. The solution is supposed to consist in saying that there are only community relative conditions (not truth-conditions) under which it can be asserted and settled whether someone is following some rule *R* as opposed to some rule *R´*.

---

[335] We are formulating this in the way in which the skeptical paradox would've put it originally. We now know, given Kripke's anti-functionalist argument, that (1) is true only if all possible determinants of the rule a person follows are physical properties of the person.

[336] Searle and others have pointed out that the very same considerations that lead to (1) can be applied to the community so as to make (2) true. So the skeptical solution does not really solve the skeptical paradox; on the contrary, it is equally subject to it.

[337] Kripke defined the concept of 'forms of life' as "the set of responses in which we agree, and the way they interweave with our activities" (Kripke, 1982, p. 96) There is an extended, metaphorical use of 'forms of life' which Kripke explores in discussing Chomsky's views. "In particular, according to Chomsky, highly species-specific constraints – a 'form of life' – lead a child to project, on the basis of exposure to a limited corpus of sentences, a variety of new sentences for new situations. There is no a priori inevitability in the child's going on in the way he does, other than that this is what the species does." (Kripke, 1982, p. 97 n. 77)

Oddly enough, when considering the skeptical paradox at the level of the individual or individual psychology, Chomsky does not seem attentive enough to the fact that it is (1) not (2) where the fundamental threat to his program lies. The skeptical paradox implies (1), not (2). Even if you do away with (2), you do not thereby do away with (1). To wit,

> […] if I follow *R*, I do so without reasons. I am just so constituted. So far, these conclusions offer no serious challenge to the account discussed earlier [the principles and parameter account of language]. I follow *R* because S_0 maps the data presented into S_L, which incorporates *R*; then 'I apply the rule blindly.' There is no answer to Wittgenstein's skeptic and there need be none. My knowledge, in this instance, is ungrounded. I know that 27 + 5 = 32, that this thing is a desk, that in a certain sentence the pronoun cannot be referentially dependent on a certain noun phrase, and so on, as a consequence of knowing rules, which I follow … I have no grounds for my knowledge in any useful general sense of the terms and no reasons for following rules: I just do it.
>
> (Chomsky, 1986, p. 225)

But this is to miss the point of the argument and to beg the question. The point is not that I follow rules but do so with no grounding or justification.[338] The point is, rather, that I do not follow any rule at all—at least not on account of any properties or features of mine. If there is no such thing as following a rule, there is no such thing as the grounds or justification for following a rule. Chomsky is more than happy to say that we follow rules blindly. But which rule or rules? Suppose Chomsky responds '*R*'. Then the question is, on account of what? What determines that it is *R* and not some other incompossible rule that is being applied? Any fact, property, or feature of the individual he appeals to can be made out to accord with some other incompossible rule and so will satisfy incompossibles, says the skeptic. But then there is no fact of the matter as to which rule, and if any, is realized.[339] This is the skeptic's argument.

For some reason the crux of the skeptic's argument escapes Chomsky. He does not think the skeptical paradox at the level of the individual or individual psychology poses a considerable problem for his account. He takes issue instead with (2). But, of course, what (1) says is that I do not and cannot follow a rule at all—at least not on account of some property, capacity, etc., of *mine*. One thing is to say that one can follow a rule without any rule-understanding (something which Chomsky accepts as possible, and Wittgenstein repudiates as nonsense), another thing is to say that one cannot follow a rule at all, with or without understanding (ignorantly?), except in an extrinsic and observer-dependent way. Chomsky seems oblivious to the fact that it is the latter one that the skeptical paradox purports to prove. Of course, whether rule-following requires or presupposes rule-

---

338 Though in part, yes. Kripke does mention that if the skeptic is right it seems that any answer we give to a mathematical problem is a leap in the dark, an instance of following a rule blindly. But to follow a rule blindly, without a definite standard of correctness, is precisely to follow or apply no rule at all insofar as rules are normative (at least the ones under consideration, i.e. formal rules), that is, standards of correctness for their applications or use. Now, Chomsky wants to hold that one can follow rules in our sense blindly, for one need not be conscious or aware of the standard of correctness one is or needs to adhere to. Wittgenstein would repudiate this, and so have we (chapter 1), but this is neither here nor there. For the point I want to press home is that the skeptic's argument entails that there cannot be such a thing a following a rule, blindly or otherwise.

339 I am formulating the argument in terms of rules, not in terms of principles and parameters, since Chomsky himself considers it in terms of rules. The two are interconnected but are generally differentiated by Chomsky and Chomskyan linguists. UG models before the 1980's are considered to be rule-based models (or models that consider language and universal grammar as a system of rules) whereas UG-models after the 1980's are considered to be principle and parameter (P&P) models where rules or systems of rules are posited as derivations from triggering of the principles and parameters rather than part of universal grammar.

understanding is an important conceptual issue. Chomsky, who posits innate formal rules or principles that are somehow followed and applied unconsciously, is committed to the view that rule-following is possible without rule-understanding. For the moment, I will leave aside issues concerning the intelligibility of this view (see chapter 1 for my remarks on the internal connection between rule-following and rule-understanding). What I want to point out here is that if (1) is true, UG collapses, and (2) is wholly irrelevant since UG posits rules and principles as innate to the individual not the community and does so on account of the sort of biological endowment of the individuals of the species in question (*homo sapiens*). This is why the rules or principles of UG are called *universals*: each and every human person is born with the universal grammar installed in the mind/brain component and with the innate ability to follow and apply these abstract formal rules or to adhere to these principles and carry out the relevant operations on the parameters. But if it's indeterminate which rule or principle-parameter structure a given person follows, applies, or has, then it is indeterminate which rules or principles are innate and therefore whether there is any fact of the matter about universal grammar at all. Kripkestein's solution to the skeptic's paradox is precisely that there is no observer or interpretation independent fact about whether someone follows a given formal rule and that rule-attributions are non-factual and devoid of truth-conditions. So, Chomsky needs to reject (1), not just (2).

Thus far we have argued that the problem for Chomsky's account does not arise only in regard to Kripkestein's solution to the skeptical paradox, that is, in regard to what justifies and constitutes *rule attributions* (e.g. how can I tell or how am I justified in saying that you are following formal rule *R* and not formal rule *R´*) but rather in regard to the considerations that lead to the skeptical paradox, that is, in regard to whether there is a fact of the matter (e.g. a set of properties) about the individual rule-follower in virtue of which it follows (or can follow) a definite rule. Chomsky's response is an *ignoratio elenchi*. This is further exemplified in his reading of (1) as a non-sequitur, a proposition fallaciously inferred on the basis of an unwarranted assumption about an internal connection between rule-understanding and rule-following. In *PI* §202 Wittgenstein writes,

> …to think one is obeying a rule is not to obey a rule. Hence, it is not possible to obey a rule 'privately'; otherwise, thinking one was obeying a rule would be the same thing as obeying it.

Chomsky agrees with the first statement of this passage but takes issue with the conclusion, considering it a non-sequitur. Of course, he says, it is true that to think one is obeying a rule is not to obey a rule. But one can obey a rule without thinking one is obeying a rule and therefore it is possible to follow rules privately. "If we say that Jones is obeying a rule "privately," nothing follows at all about whether he thinks that he is obeying the rule. In fact, we do say that Jones obeys a rule "privately"… even if he thinks he is obeying a different rule or has no idea about rule following (and is responding differently than we would)" (Chomsky, 1986, pp. 229–230). Put in context, however, one can see that Wittgenstein gives independent considerations as to why one cannot follow a rule without knowing what it is to follow the given rule, as to say otherwise is to speak nonsense. Chomsky, on the contrary, is precisely of the opinion that one can follow a rule without knowing one is following the rule or what it is to follow the given rule (indeed, without knowing any rule at all). As a response to the skeptical paradox and (1), Chomsky's rejoinder simply misses the point.[340]

---

[340] Chomsky follows this up with the cautionary advice "we must avoid the temptation to assume some notion of 'accessibility to consciousness' with regard to mental states and their contents. This seems inconsistent with the way we use the relevant concepts in normal discourse, however deeply rooted the assumption may be in various philosophical theories, and it is an assumption that can only stand in the way of a descriptively adequate or genuinely explanatory

As a response to whether rule-following requires rule-understanding, one could argue that it is misplaced and misconceived. But this not what I want to highlight here. We have already given reasons as to why one cannot be said to follow a formal rule if one does not understand the formal rule in question (see chapter 1).[341] We can, for the sake of argument, concede that conceptual issue to Chomsky. The crucial point lies, rather, elsewhere: namely, in Chomsky's inability to see that the skeptical paradox implies that there is no such thing *in the individual* that determines which *formal* rules or structures she follows or is endowed with. There is no fact of the matter. Anything that could be said to confirm formal rule R and therefore some UG theory *T* will be said to confirm an incompossible formal rule R´ and therefore some other incompossible UG theory, *T´*.

Now, I do not want to refute UG here. It is not the aim of this chapter to purport to carry out such a task. What I want to focus on is on Chomsky's views about the origin and possible determinants of the universal grammar innately found in the mind/brain of the individual human person and the implications KRA has for his answer to the genetic problem. For even if we grant that there are a set of formal structures and rules innate to the 'language faculty' of the human being, we cannot intelligibly predicate these innate formal rules of any material aspect or part (e.g. the brain) of the human being. In light of KRA, we now know that (1) is incoherent but

(1´) There is no physical-material fact(s) *about someone* that can determine whether she is following a formal rule (that is, whether she is following *R* and not *R´*).

is not. Chomsky, however, rejects (1´). According to the Chomsky, the aspects of linguistic understanding and knowledge (namely, the formal structures that constitute UG) that are not and cannot be learned from experience and do not come from without but from within, are innate to the psyche of the human person. But the innate grammatical structure in the psyche of the individual person is neurobiologically determined, part of our genetic inheritance, although concessions are made to the effect that at the moment we are not able and may not ever be able to account precisely for how this is so (the genetic problem). Chomsky is known for having posited a 'language organ', an innate module or faculty contained in the brain of human beings that comes with built-in universal grammatical rules, or principles and parameters. As we've already pointed out, these principles and parameters are *tacitly known* (*cognized*), that is, unconsciously known, and the relevant rules and functions are applied and computed unconsciously. Chomsky wants to hold that these innate aspects are evolutionarily grounded and neurobiologically determined, even if the sciences are too far away from achieving a unified conception of the human psyche. He holds this to be an intelligible claim. But if our previous chapters are on the right, some of the implications KRA inevitably undermine Chomsky's neurobiologism about the 'language faculty'.[342] His claim that the

---

theory of mind." (Chomsky, 1986, p. 230) But the claim that rule-following is dependent, and consequent upon, rule-understanding assumes no such notion of 'accessibility to consciousness', and thus just betrays Chosmky's misreading of Wittgenstein's notion of understanding as some sort of introspective process whereby a rule is hit upon. To be fair, Kripke makes a similar mistake but rightly rejects such 'accessibility to consciousness' as insufficient for determining a definite formal rule.

[341] For a very good critique of Chomsky's conception of rule following, see (Hacker, 1990). In my estimation, one has to conclude that Chomsky does not understand what a rule, and in particular a formal rule, is. Maybe the rules and principles of UG are meant to be *sui generis* but then it is not clear in what sense they are rules or principles.

[342] Again, I am assuming for the sake of argument that there is such a thing as a language faculty or language organ. "This language organ, or "faculty of language" as we may call it, is a common human possession, varying little across the species as far as we know, apart from very serious pathology. Through maturation and interaction with the environment, the common language faculty assumes one or another state, apparently stabilizing in several stages, finally at about puberty. A state attained by this faculty resembles what is called "a language" in ordinary usage, but only partially: we are no longer surprised when notions of common sense find no place in the effort to understand and explain the

universal formal rules or principle-parameter structure are contained in the brain and were biologically (e.g. evolutionarily) determined needs serious reconsideration. As a dialectical concession, I will talk of formal capacities, abilities, and activities instead of 'formal understanding' so that no one can accuse us of begging the question against Chomsky, for whom the formal structure of the language faculty that constitutes competence is tacitly known and therefore 'tacitly understood', if that expression even makes sense. In our previous chapters, we have already given a partial answer to the constitutive question about our 'formal' capacities, abilities, and activities. I will ignore Chomsky's claim that some linguistic *abilities* (as opposed to capacities) are not acquired but innate.[343] Let us simply rehash our argument in terms of formal capacities, abilities, and activities and explore some of the potential problems KRA poses for Chomsky's conjunction.

Chomsky takes himself to have identified two different senses of 'forms of life' in Kripke's skeptical argument. The first sense is the *narrow* sense: forms of life refers to the set of responses a given community agrees in and the way they interweave with the community's activities (or, rather, the activities of the members of the community). The second sense is the *wide* sense: it refers to species-specific constraints that somehow set the conditions for the possibility for rule-following in the narrow sense. Now, Chomsky takes the use of 'forms of life' in the narrow sense to be at the level of a particular grammar, i.e. the attained language (e.g. Spanish, English, Dutch, Latin, etc.). The use of 'forms of life' in the wide sense is, according to Chomsky, at the level of UG (the initial state, represented as 'S_0') (Chomsky, 1986, p. 232).[344] Chomsky uses this distinction to argue that, in the wide sense, there is no private language but rather universally shared linguistic structures but that in the narrow sense, there could be a private language and therefore Wittgenstein's private language argument collapses. This betrays a misunderstanding of Wittgenstein's private language argument. But this is neither here nor there. Our point is that if there is such a thing as 'forms of life' in the wide sense and therefore such a thing as UG, then UG cannot be (wholly) material and cannot be fully accounted for constitutively and genetically in terms of the neurobiology and evolution of human beings. The brain and biological constitution of a human being is just as formally indeterminate as any other physical mechanism. In a way one can say we have saved the individual psychology framework of generative grammar from the skeptical paradox by showing that (1) is incoherent. For if the capacities and abilities of a human person were nothing more than material capacities biologically determined, then no formal rules or principles could be determinately followed and applied by a human person. There would not be, because there could not be, such a thing a rule-governed activity in regard to formal rules or such a thing as formal (e.g. computational) operations. Consequently, there could be no such thing as UG. But since human beings do carry out formal operations and apply formal rules, since they do partake in rule-governed activity that is formally determinate, then the relevant capacities, abilities, and activities cannot be wholly a matter

---

phenomena they deal with in their own ways, another achievement of the Galilean revolution, now taken for granted in the hard sciences but still considered controversial beyond – inappropriately, I think." (Chomsky, 2002, p. 47)

[343] I leave it open whether it is logically possible to be born with or to come to have (say, by a miracle) a set of abilities which one did not acquire. This might be 'logically' possible, whatever that means (i.e. conceivable?).

[344] The thrust of the hypothesis is this: the initial cognitive state of a human being (that is, the cognitive state at birth) rather than being a *tabula rasa* in the way empiricists would have it, is instead a richly structured state with principles, parameters, and rules or functions. This is the rationalist or Cartesian aspect of Chomsky's project. This cognitive state (or, rather, the structure of this initial cognitive state) is the innate universal grammar. Language acquisition is then seen as the complex transition (made possible by the innate, tacit, and unknown structure of S_0 in combination with exposure to a particular language in the course of experience) from the initial cognitive state, S_0, to the stable state, S_L, that corresponds to a particular language, e.g. English. This particular stable state S_L is what the theoretical (generative) linguist calls 'a particular grammar' and refers to the knowledge (competence) that a person has of a particular language (say, English or Spanish). The theoretical linguist's task is to investigate, explain, and make hypotheses about the formal structure of S_0, i.e. universal grammar.

of matter. So UG can, of course, survive. But only if it rejects its neurobiologism about formal abilities and activities.[345] Yet this would mean renouncing the claim that "the statements of grammar and UG are not different in principle from the statements of natural science theories; they are factual, in whatever sense statements about valence or chemical structure or visual processing mechanisms are factual and involve truth claims." (Chomsky, 1986, p. 224) They might be truth-claims or 'factual', but not in the same sense in which statements about material structures are.

So, if natural languages involve formal operations (say, the computation of recursive generative functions[346]), then natural languages (or, rather, the capacity for natural languages) cannot be constitutively and genetically (in our sense of those terms, see chapter 1) wholly a material phenomenon and neither can linguistic understanding and knowledge which requires the ability to carry out the relevant formal operations. But a concession needs to be made. If not for the kinds of animals we are (e.g. our anatomy, our neurobiological mechanisms, etc.) there could be no such thing as human languages as we know them. Our species-specific endowment together with our environment are a necessary condition for the possibility of human languages and may even be sufficient for certain aspects of languages (e.g. phonetic sounds, constraints on the length of words, etc.), but they are not and cannot be a sufficient for the logico-grammatical structures of languages and linguistic understanding. This is our point. Now, one does not need to hold to the existence of a universal grammar to hold that natural languages are rule-governed and consist of logico-grammatical structures, many of which are correctly characterized as formal and which the language speaker follows and applies. The result would be the same: insofar as the logico-grammatical, syntactical rules and structures of natural languages are formal (that is, have a pure abstract invariable component), they cannot be wholly explicable in material terms. So the conclusion would be the same. But there are two main points of contention which bring us back to Chomsky's conjunction. First, whether there is any such thing as universal grammar in either the brain or the mind that is composed, in part, of formally determinate rules, functions, principles, and mechanisms that are innate. Second, whether these innate formally determinate rules, functions, principles, or mechanisms can have as a sufficient condition for their formal determination a given material (e.g. biological) endowment. We have not concerned ourselves with the first statement, even though we've expressed our reservations. For all we've said, the nativist's hypothesis of a universal grammar might be true. But we have railed against Chomsky's position regarding the second statement and pointed out that, in light of KRA, it cannot be true. Therefore, if UG is correct about our linguistic capacities, their innate structure, and their role in language acquisition, it cannot be so simply on account of what Chomsky would want to assert: the sufficiency of material conditions of some kind for determining the innate formal system of universal grammar.

4. CONCLUDING REMARKS

In this concluding chapter, we've explored some of the consequences KRA has for Chomsky's conjunction and therefore for the philosophical presuppositions he brings to bear on the Central

---

[345] Do not confuse neurobiologism about formal understanding with the neurobiology of formal understanding and of any kind understanding, which is a *necessary* condition and grounding of our cognitive abilities and activities. The latter points to one of the species-specific necessary conditions for *human* formal understanding and for humans acquiring or having the ability to follow and apply formal rules. I take this to be incontrovertible. The former, on the contrary, points to the *sufficiency* of our species-specific neurobiological constitution for determining a definite, unique set of (innate) formal rules, principles, and parameters. This is not possible, as KRA shows.

[346] This is the underlying and consistent thread throughout all of the different models of UG that Chomsky has proposed through the span of five decades, from the original model of UG to the latest, the so-called Minimalist Model (Chomsky, 1995c, 1995a, 2000).

Project of theoretical linguistics. There is a conceptual tension between (a) his claim that universal grammar is, in part, the formally rich structure of the language faculty and (b) his claim that our species-specific neurobiological is sufficient to determine the formal components of the innate structure (of the universal grammar) of the language faculty. We have pointed out that Chomsky cannot consistently hold on to both, and that if he wants to hold on to UG as he conceives it he must renounce (b). An alternative is to abandon Cartesian linguistics altogether without renouncing the formal structure of natural languages and linguistic understanding, and to develop a conceptual, ontological, and metaphysical framework that can explain human behavior as intelligible and transparent as it relates to linguistic, mathematical, and logical activity. 'Performance', as Chomsky understands it, as the "physical manifestations of knowledge of language" disappears and linguistic activity with its formal and semantic structures becomes readily intelligible and perceptible to beings of a certain sort. That's one option. Chomsky's conception of 'behavior' is a crucially impoverished one that gives rise to *pld* considerations and arguments from the poverty of stimulus. Though this is well beyond the scope and resources of the present project, it is important to note that one need not be committed to a Cartesian view of human behavior (for alternatives, see e.g. Braine, 2014, 1992; Hacker, 2013, 2007; Hacker and Baker, 2014, 2005; MacIntyre, 1986, 2006, pp. 86–103). Our treatment of Chomsky's philosophical commitments has not been exhaustive, and we do not pretend to have refuted or definitively argued against them. We've merely posed in a preliminary manner a dilemma that we believe KRA makes Chomsky subject to. A deeper and more complete treatment of this complicated issue would involve, for example, a detailed examination of the different paradigms of generative grammar proposed by Chomsky, an exploration of the main similarities and differences, an articulation of the nature of the technical machinery introduced in search of explanatory and descriptive adequacy, and showcasing in greater detail the perennial thread underlying all of the different paradigms that makes them, in light of the architectural role Chomsky's conjunction plays in the conceptual underpinnings of the Central Project of theoretical linguistics, subject to our kind of argument. This must await not only another occasion, but another monograph.

A note on the scope of our argument. We've explored some of the difficulties Chomsky's approach to language faces if the relevant structure that they purport to explain is both formal and neurobiologically determined. We've extended the implications of KRA to the capacity for language on the assumption that the capacity for language is a formally determinable (or formally determinate) capacity. If it isn't, then language will not fall within the purview of KRA, though formal (e.g. mathematical or logical) understanding will. But then neither would language fall under Chomskyan approaches, explanatorily speaking. If the capacity for language, whether in potency (mere capacity) or in actuality (ability), is not formally determinable then Chomskyan approaches if they are to be saved cannot be said to uncover or discover, through formal and abstract models, the real nature of universal grammar. Chomskyan approaches would be formalizations of a non-formal reality, and therefore subject to the indeterminacy and approximation considerations presented in this thesis. We have also, for dialectical purposes, made no use of the claim that the ability to carry out any given formal operation is consequent on, or is constituted by, understanding the formal operation in question at the relevant level or degree of competence (in our sense, see chapter 1). Chomsky may not grant that there is understanding at the level of universal grammar but he surely must grant it even if not in *toto* at the level of particular grammars (like English, Spanish, Chinese, etc) in *competent* speaker-hearers. Perhaps full mastery of particular grammar requires a tacit and unknown set of abilities that are unconsciously exercised in mastering a particular natural language. But that mastery of a particular natural language must involve language specific understanding is something Chomsky does, of course, concede. That's all I need. We can make the argument both at the level of universal grammar and at the level of particular grammars (i.e. natural languages), or at either one. Our case

will remain regardless of its scope. In the case of particular grammars, there is no mastery of a specific natural language without some sort of understanding of the specific language and its grammatical rules (that is, grammatical normativity), since to master a natural language is, partly, to come to understand the grammar of a particular language and to acquire a set of language-specific abilities with respect to specific grammatical normativity of the language in question (see, for example, our discussion of competent judgements in §1.2 and §1.3). It therefore will consist in an array of abilities that are and can be variously manifested in speech, speech comprehension, intelligible verbal action, and so forth, or more simply, as Chomsky often emphasizes, in judgments about the grammaticality of utterances. Whether it involves more than that (say, a universal grammar) is an issue that goes beyond the aims of this chapter. But even if, for the sake of argument, there is no understanding at the level of mastering a particular grammar (something hard to believe), there is still the activity (and ability) of carrying out highly abstract formal operations, something which Chomsky claims humans, and more peculiarly children, do (albeit unconsciously) when mastering a particular grammar. KRA will imply that this activity cannot have, as a sufficient condition, a wholly neurobiological locus and cause.

Finally, let me note that KRA has consequences for any approach or project in linguistics insofar as it raises the question of whether language is constituted by structures generated by formal rules of some sort or whether it is a condition for the possibility of grasping, thinking, following, pure formal structures and rules at all. KRA bears on the nature of our capacity for language and on the nature of the actualization of that capacity in the acquisition of a language and of linguistic understanding. Merely evolutionary approaches to languages will fail for similar reasons if language involves formal structures or formal operations, or if language is that which makes formal understanding and operations possible. Evolutionary approaches to the natural linguistic *capacity* of human beings have *in principle* explanatory limits if understanding a natural language involves formal understanding. But also if linguistic understanding is a condition for the possibility of formal understanding. Language is an incredibly complex phenomenon, and there is both room for formal and non-formal approaches. Chomsky's Cartesian divide between I-language and E-languages, and strict separation between competence and performance, need not be accepted in order to admit that languages have a formal structure and that formal tools for the study of language help us uncover or understand better the nature of language in certain respects. Sociolinguistics and particularly the field of pragmatics have as much to say about the nature of language as formal approaches do. There have been some attempts (for an excellent and incredibly well informed Aristotelian-Wittgensteinian approach to language and linguistics, see Braine, 2014) to give a more holistic view of language and linguistic understanding that rids itself of any Cartesian model (be it substance dualist and rationalist or the reactions to these, like physicalism and empiricism). Maybe as Nagel pointed out (Nagel, 2012), we need to reconsider some sort of Aristotelian approach to the cogitative powers of human beings that recognizes capacities and abilities that belong to the individual members of the species but which does not ignore the environment nor the fact that human activity is, to a great extent, practice-constituted activity and must be situated at a broader level than the individual. Hence, the Wittgensteinian side of this approach. This, I believe, is a project that must be carried out further if we wish to arrive at a more adequate framework for explaining the nature of linguistic understanding and practices as well as the capacity we have for such understanding and practice. But as far as the aims of this chapter and monograph go, the formal aspects of our cogitative powers will not be explicable, both constitutively and genetically, if the possible explanations of such powers require the sufficiency of material conditions in one way or another. If language is a species of formal understanding or a condition for the possibility of formal understanding, this much will be true of our linguistic capacities too.

CONCLUSION

In the course of this monograph we have argued for a host of different claims, the main one being that there are aspects of human understanding that cannot be explained, constitutively and genetically, in wholly materialist terms because they are not materially or physically determined. In chapter 1 we laid down the conceptual framework for the whole work, distinguishing between a capacity, an ability, and an activity, and explaining what was meant by the concepts of 'formal understanding', 'pure formal structures', and the like. In chapter 2 we gave our Kripkean argument for the formal indeterminacy of the physical and showed how accounts of physical computation could not provide a successful explanation of formal understanding and activity. We then addressed, in chapter 3, a whole range of objections against the argument we laid out in chapter 2 and found them all wanting. In chapter 4 we went on to argue for the formal determinacy of formal understanding and formal activity. In that chapter we also developed Ross's own considerations regarding the conceptual perils of denying that human understanding could have any formal determination intrinsically. In the following chapter (chapter 5) we then addressed some of the main objections that could be raised against our Rossian argument for the claim that what is true of the physical with respect to formal rules and operations cannot be not true of the cogitative (i.e. of formal understanding). In chapter 6 (which reads more like an appendix) we extended, albeit in a provisional and propaedeutic manner, the line of reasoning developed in the preceding chapters to Chomsky's computational and formal approach to the nature of the capacity for language, linguistic understanding, and natural languages. There we showed that if Chomsky is correct in the sense that language (or the capacity for language) is, partly, formally structured then linguistic understanding and the capacity for such understanding cannot be, constitutively and genetically, wholly a matter of matter. Similarly, we made the point that if linguistic understanding is the condition for the possibility of formal (mathematical and logical) understanding, whether as a by-product or as an abstraction from the formal features of language, then the capacity for language cannot be naturalized. Much more needs to be said, however. In this conclusion I lay out some of the most pressing problems that are left untouched in the present work and that merit further investigation, particularly if we wish to assess the overall philosophical cogency of KRA.

The first major issue concerns an account of the mathematization of nature (which, for example, physics has carried out) and the successes in the application of mathematics for the manipulation of matter (e.g. in engineering, computer science, nanotechnology, etc.). This requires an analysis of the intellectual means by which the material is ordered, studied, explained, and manipulated *using* formal tools and constructions and we must explain how is it that we can order, explain, study, and manipulate matter by reference to and in virtue of any formal structure at all (like mathematical equations) given what KRA implies, namely that matter can only realize formal structures imperfectly, extrinsically, and relative to us. This requires an investigation into the philosophy of mathematics and the philosophy of science which would have been impractical and unfeasible given reasonable limitations imposed on the present work. Not all can be done in a single monograph. Yet this is an important problem that must be addressed.

The second major issue concerns the possibility of our capacity for languages (bot natural *and* artificial languages) and linguistic understanding. Some certain strands in linguistics attribute to human beings a language faculty that is *formally* rich and structured, one which counts as the condition for the possibility of acquiring any particular language (linguistic competence). But if this is true, then our argument will apply to linguistic understanding too and therefore the human capacity for language cannot be amenable to a purely materialist account. In chapter 6 I sketched out some of the consequences that KRA has for formal approaches (i.e. Chomsky's) to linguistic

understanding. But much more needs to be said, and further research on this area needs to be carried out. If, however, linguistic understanding is unlike formal (i.e. mathematical and logical) understanding, it still remains that language is the *sine qua non* condition for coming to understand anything mathematical and logical at all since *understanding* is, first a foremost, a linguistic phenomenon. It is made possible by language. So our argument may still establish the impossibility of naturalizing the capacity for language insofar as that capacity undergirds all manners of understanding, including mathematical and logical understanding. This would also help explain why only linguistic animals (animals with the capacity for language) are the only ones capable of mathematics, logic, physics, engineering, architecture, programming, economics, and so forth. Moreover, in this monograph, I situated formal understanding within *practices* yet did not, due to reasons of space and scope, address the nature of practices and the particularities of the complex conceptual web that is part and parcel of *formal* practices, particularly of refined and well-established formal practices like mathematics and logic.

The third major issue is related to the second one. Because formal understanding is acquired and manifested in practices of a certain sort, there must be formal intelligibility to human action embedded in formal practices. The non-naturalizable aspects of formal understanding and of formal activity must somehow be perceptible to other human beings who possess the relevant kind of understanding. Somehow who has no knowledge of linear algebra, will be unable to perceive and understand what a visiting scholar is proving on a black board. We must explain the possibility of *teaching* mathematics and formal logic, of *manifesting* or *exhibiting* excellence and mastery in mathematics and logic, of *communicating* to other human beings formal understanding, of the *shared* or *public* standards of correctness of such practices, and so forth. So the direct intelligibility of human action as embedded in such practices is also something that must be accounted for, and if I am right about the nature of language, then the direct intelligibility of speech (as the exercise of linguistic understanding) must also be accounted for in a non-naturalistic way. Much of this will be an interdisciplinary oriented approach, but the conceptual elucidations regarding formal determinacy if correct must be kept in sight when attempting to explain formal activities and practices.

Formal understanding, we have said, is acquired and its acquisition is generally the outcome of learning by teaching and training ('self-taught' counts). That is to say, we *learn* (to do) mathematics, we learn (to do) logic, we learn (to speak and understand) a language, and so on. We don't learn these things in the exact same way, but that we do learn them (or some of them) is a fact. The question that now presents itself is how exactly is this possible *given* KRA. The problem is a familiar one: if formal structures are outside of space and time, eternal or sempiternal, and so forth, then how can we (rational *animals*) have any 'contact' with them? How can we teach and learn them? Putting to the side the misleading expression 'coming into contact with', it seems that behavior is an essential aspect of teaching and learning. But our behavior, including linguistic behavior, is embodied and therefore *qua* physical unable to be structured in any formally determinate way. But then it seems that there would be no way of communicating and conveying determinately, short of telekinesis, the proper objects of formal understanding. A teacher instructing or teaching her pupils to add or to borrow and carry could just as well be said to be instructing or teaching her pupils to quadd or to 'qu-borrow and qu-carry'.

There are several conceptual problems intertwined here and clarity on the relevant concepts is of paramount importance. Therefore, this worry cannot be answered until the concepts of 'behavior,' 'teaching,' 'learning,' 'rule,' etc., are clarified. Secondly, this is a problem not peculiar to KRA, but to any philosophical standpoint that acknowledges pure formal structures, whether as subsistent entities in a third realm, or as abstractions from real, natural, structures, or as mere arbitrary but useful constructs and fictions. The other option is to deny that there are any pure formal structures and/or to say that pure formal structures are really natural or material structures, such that matter

does make it into the definition and ontological identity of such structures. But our argument shows that this cannot be an intelligible option. What needs to be explored, instead, is the precise nature, the metaphysics, of pure formal structures as such. One thing is to say that these kinds of structures exist, another is to say *in which way* they exist (as fictions? As mere intellective entities or constructions? As Platonic entities?). Perhaps it is worthwhile to point out that it is not the case that in order for something to be communicated, even in linguistic behavior, it must be of the same ontological category as that which is used to communicate it. When a teacher writes '+2' on the board and instructs her students to complete the following series '2, 4, 6, 8, …' up to 100 according to the rule '+2', she's using physical items (e.g. the '+' and '2' written on the board or on this page) as *symbols*, that is, as physical items *interpreted* according to certain conventions about mathematical notation and representation. On account of the physical properties alone, nothing formal or semantic is conveyed or communicated by these physical items. They are in themselves semantically indeterminate. Yet there is no doubt that when a pupil, trained in arithmetic, sees '+' written on the black board she *sees* (so to speak) addition. So there is a way in which what is formally-semantically indeterminate can carry or contain something that is formally-semantically determinate, like circuits and logic gates, but it cannot be on account of any of the intrinsic material properties of the former that this comes about. Use, interpretation, convention, agreement in 'forms of life' (as Wittgenstein would say), all supply the formal-semantic determinacy to physical items; that's how they *become* symbols suitable for formal operations. Of course, some notations are better or more convenient than others and this has to do in part with their physical structure; matter *does* matter. But physical structures do not amount to, nor do they suffice to determine definite, formal and semantic structures. In order to assess the meaning of such items, one needs to know how they are *used* in the practices in which they figure as symbols (see our example of sigma notation in chapter 1). The same goes for physical computing mechanisms. From the discreet items they process and display (in the case of digital computation) to the continuous variables they process (analog computation), the mechanism must be situated in a broader context of a human practice where it can be interpreted in order to be determinate with respect to what it is doing. They do what they do by convention. In itself and objectively speaking, a physical computing mechanism satisfies incompossible pure functions or structures and therefore realizes none. But that symbols can be representational is undeniable. We use '+' to represent addition, and '2' to represent the number two. But those symbols could be used, and are in fact used, to represent altogether different rules and "entities"; there is nothing in the properties of '+', for example, to stop it from representing quaddition. Convention is where the buck stops. A study of the nature of conventions, particularly formal conventions, is needed if we wish to provide a more complete explanatory framework of formal activity and formal representation.

Another problem that KRA seems to raise has to do with whether its conclusion is true of any and every mathematical activity, or whether it applies merely to a subset of such activity. For example, is all of this true of even ordinary folks when they are calculating the cost of their weekly groceries, or is this rather true of the more specialized, abstract skills of mathematicians, logicians, and scientists? Or, perhaps, the scope is even more restricted: does it really apply to what mathematicians do (that is, to their understanding and activities) or rather to the regimentation and idealization, via formal systems, of what they do? If the latter, it would still remain that KRA would be true of the ability to regiment and idealize formally the types of mathematical and logical rules and operations that mathematicians routinely apply and carry out. As Peter Smith puts it,

> … there are formal systems including the so-called ω-rule, which says that from the infinite array of premisses $\varphi(0)$, $\varphi(1)$, $\varphi(2)$, . . . , $\varphi(n)$, [. . .] we can infer $\forall x\varphi(x)$

> where the quantifier runs over all natural numbers. But do note that finite minds can't really take in the infinite number of separate premisses in an application of the ω-rule: that's an impossible task. Hence, *in so far as the business of formalization is primarily concerned to regiment and formalize the practices of ordinary mathematicians, albeit in an idealized way*, it's natural at least to start by restricting ourselves to finite proofs, even if we don't put any particular bound on the length of proofs.
>
> (Smith, 2013, p. 30 n. 8; my emphasis)

Now, we may wonder whether formalized systems of logic and mathematics aren't but idealizations of actual logical and mathematical activity, that is, of what logicians and mathematicians do and of what ordinary people do when they do logic or do mathematics. This is an interesting observation but affects only the scope not the substance of KRA. For, who does and carries out the regimentation, the formalization, and the idealization of logical and mathematical thinking? We do. These are formally determinate activities we can engage in and about which we carry out mathematical or logical proofs (e.g. Gödel's completeness and incompleteness theorems). So, the capacity for such activity will be subject to the results of KRA. Secondly, even if we concede for the sake of argument that KRA applies only to the sort of practice and activity which mathematicians and logicians (that is, specialized experts on mathematics and logic) engage in and not to the ordinary, everyday practice of non-expert mathematical and logical thinking (say, calculating as you go how much you have in your grocery bag to make sure you stay within your budget or arguing validly or invalidly for the best economic policy), it would still remain that any sort of being that is capable of becoming a mathematician or a logician (or both) has a capacity that cannot be, constitutively and genetically, wholly a matter of matter. For the capacity to become a mathematician or a logician is the capacity to engage in formally determinate activity, an activity that cannot be explicable in physical or material terms alone, be them reductive or non-reductive. Since it is true that human beings are the sorts of beings who have that cogitative capacity, it follows that human beings have capacities and powers not wholly accountable and attributable to matter. This is a *metaphysical* statement, since it is true of every human being that they have such a capacity. But in practice, that capacity is not always developed, and, in fact, it need not be developed (and in cases of disability and impairment, the development of that capacity is actually hindered). So not every human being has the *ability* to do mathematics and logic at a suitably complex level, let alone at the level of the mathematician and logician. Yet natural languages may aid in the development of that ability, even if at a rudimentary level, by containing grammatical rules that allow for argumentation and for quantification and calculation (say, counting and subtracting). An interesting and fascinating investigation would be the interconnections between the capacity for linguistic understanding and the capacity for formal understanding. As far as we can tell, only beings capable of language are capable of formal understanding and developing formal practices, which is why science and philosophy for example are a uniquely human phenomenon in the animal kingdom.

All in all, many issues remain unresolved and unanswered. Yet the primary research objective of the present work has, to our estimation, been achieved. We have argued extensively for the soundness of KRA and take ourselves to have answered successfully some of the most pressing objections facing the argument. It is impossible to foresee all of the implications of and objections to KRA, and further attention from specialists in areas whose subject matter overlaps with the claims of KRA can illuminate possible faults in the argument as well as possible ways of improving it, assuming any possible fault that might be found does not turn out to be fatal. At any rate, we do not claim to have exhausted the potentialities of KRA nor to have given the best possible variant and defense of that argument. Maybe KRA will not survive further scrutiny. Or maybe it will. This

remains to be seen. But regardless of the merits of the argument, KRA raises important issues about the nature of our capacity for formal understanding and activity, of our ability to do such things as mathematics and logic and to apply or follow formal rules in a variety of ways and for a variety of purposes, and about the proper ways in which such phenomena should be explained. Such issues are of paramount importance if we hope to better understand the nature of human understanding.

BIBLIOGRAPHY

Abelson, H., Sussman, G.J., Sussman, J., 1996. Structure and interpretation of computer programs, 2. ed. ed, Electrical engineering and computer science series. MIT Press [u.a.], Cambridge, Mass.

Adger, D., 2003. Core syntax: a minimalist approach, Core linguistics. Oxford University Press, Oxford ; New York.

Ahmed, A., 2007. Saul Kripke, Continuum contemporary American thinkers. Continuum, London; New York.

Anscombe, G.E.M., 1985. Wittgenstein on Rules and Private Language. Ethics 95, 342–352.

Armstrong, D.M., 1993. A materialist theory of the mind, Rev. ed. ed, International library of philosophy. Routledge, London ; New York.

Aydede, M., 1997. Language of Thought: The Connectionist Contribution. Minds and Machines 7, 57–101. https://doi.org/10.1023/A:1008203301671

Ayoun, D., 2003. Parameter setting in language acquisition. Continuum, London ; New York.

Balaguer, M., 2018. Fictionalism in the Philosophy of Mathematics, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Balaguer, M., 2016. Platonism in Metaphysics, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Balaguer, M., 2001. Platonism and anti-platonism in mathematics. Oxford Univ. Press, Oxford.

Balaguer, M., 1995. A Platonist Epistemology. Synthese 103, 303–325.

Bays, T., 2014. Skolem's Paradox, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Beakley, B., Ludlow, P. (Eds.), 2006. The Philosophy of Mind: Classical Problems/Contemporary Issues, 2nd ed. ed. MIT Press, Cambridge, Mass.

Belletti, A., Rizzi, L., Chomsky, N. (Eds.), 2002. Editors' introduction: some concepts and issues in linguistic theory, in: On Nature and Language. Cambridge University Press, Cambridge, pp. 1–44. https://doi.org/10.1017/CBO9780511613876.002

Benacerraf, P., 1998. What Mathematical Truth Could Not Be–1, in: Schirn, M. (Ed.), The Philosophy of Mathematics Today. Oxford Univ. Press, Oxford, pp. 33–75.

Benacerraf, P., 1996. Recantation or any old w-sequence would do after all. Philosophia Mathematica 4, 184–189.

Benacerraf, P., 1973. Mathematical truth. Journal of Philosophy 70, 661–679.

Benacerraf, P., 1965. What numbers could not be. Philosophical Review 74, 47–73.

Bennett, M.R., Dennett, D.C., Hacker, P.M.S., Searle, J.R., Robinson, D.N., 2007. Neuroscience and philosophy: brain, mind, and language. Columbia Univ. Press, New York Chichester.

Bennett, M.R., Hacker, P.M.S., 2003. Philosophical foundations of neuroscience. Blackwell Pub, Malden, MA.

Bickle, J., 2016. Multiple Realizability, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Blackburn, S., 1984. The individual strikes back. Synthese 58, 281–301. https://doi.org/10.1007/BF00485244

Block, N., 1980a. Are Absent Qualia Impossible? The Philosophical Review 89, 257–274. https://doi.org/10.2307/2184650

Block, N., 1980b. Introduction: What is Functionalism?, in: Block, N. (Ed.), Readings in Philosophy of Psychology. Vol. 1: ..., The Language and Thought Series. Harvard Univ. Press, Cambridge, Mass, pp. 171–184.

Block, N., 1978. Troubles with functionalism. Minnesota Studies in the Philosophy of Science 9, 261–325.

Bloor, D., 1997. Wittgenstein, rules and institutions. Routledge, London ; New York.

Boghossian, P.A., 2005. Is Meaning Normative?, in: Nimtz, C., Beckermann, A. (Eds.), Philosophy - Science - Scientific Philosophy: Main Lectures and Colloquia of GAP.5. Fifth International Congress of the Society for Analytical Philosophy., Perspektiven der analytischen Philosophie. Mentis, Paderborn, pp. 205–218.

Boghossian, P.A., 1989. The Rule-Following Considerations. Mind, New Series 98, 507–549.

Braddon-Mitchell, D., Jackson, F., 2007. The Philosophy of Mind and Cognition: An Introduction, 2nd ed. ed. Blackwell Pub, Malden, MA.

Braine, D., 2014. Language & human understanding: the roots of creativity in speech and thought. The Catholic University of America Press, Washington D.C.

Braine, D., 1992. The human person: animal and spirit. University of Notre Dame Press, Notre Dame, Ind.

Bridge, J., 1977. Beginning Model Theory: The Completeness Theorem and Some Consequences, Oxford logic guides. Clarendon Press, Oxford.

Buechner, J., 2011. Not Even Computing Machines Can Follow Rules, in: Berger, A. (Ed.), Saul Kripke. Cambridge University Press, Cambridge, pp. 343–368. https://doi.org/10.1017/CBO9780511780622.015

Burge, T., 2007. Foundations of mind, Philosophical essays. Clarendon ; Oxford University Press, Oxford : Oxford ; New York.

Burge, T., 1979. Individualism and the mental. Midwest Studies in Philosophy 4, 73–122.

Cartwright, N., 2007. Hunting Causes and Using Them: Approaches in Philosophy and Economics. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511618758

Cartwright, N., 1999. The dappled world: a study of the boundaries of science. Cambridge University Press, Cambridge, UK : New York, NY.

Cartwright, N., 1994. Nature's capacities and their measurement, 1. issued in paperback. ed. Claredon Press, Oxford.

Cartwright, N., 1989. Capacities and abstractions, in: Kitcher, P., Salmon, W. (Eds.), Scientific Explanation. Univ of Minnesota Pr, pp. 13--349.

Cartwright, N., 1983. How the Laws of Physics Lie. Oxford University Press. https://doi.org/10.1093/0198247044.001.0001

Chalmers, D.J., 2012. The Varieties of Computation: A Reply. Journal of Cognitive Science 13, 211–248. https://doi.org/10.17791/jcs.2012.13.3.211

Chalmers, D.J., 2011. A Computational Foundation for the Study of Cognition. Journal of Cognitive Science 12, 325–359. https://doi.org/10.17791/jcs.2011.12.4.325

Chalmers, D.J., 2010. The character of consciousness, Philosophy of mind. Oxford Univ. Press, Oxford.

Chalmers, D.J., 1997. Moving forward on the problem of consciousness. Journal of Consciousness Studies 4, 3–46.

Chalmers, D.J., 1996. Does a Rock Implement Every Finite-State Automation? Synthese 108, 309–333.

Chalmers, D.J., 1995a. Facing up to the problem of consciousness. Journal of Consciousness Studies 2, 200–19.

Chalmers, D.J., 1995b. Minds, Machines, and Mathematics: A Review of Shadows of the Mind by Roger Penrose. Psyche 9.

Chalmers, D.J., 1994. On implementing a computation. Minds and Machines 4, 391–402.

Chomsky, N., 2010. Lectures on Government and Binding: the Pisa Lectures.
Chomsky, N., 2009. Cartesian linguistics: a chapter in the history of rationalist thought, 3rd ed. ed. Cambridge University Press, Cambridge ; New York.
Chomsky, N., 2005. Language and the brain, in: Saleemi, A.P., Bohn, O.-S., Gjedde, A. (Eds.), In Search of a Language for the Mind-Brain: Can the Multiple Perspectives Be Unified? Aarhus University Press ;
Chomsky, N., 2002. On nature and language. Cambridge University Press, Cambridge.
Chomsky, N., 2001. Derivation by Phase, in: Hale, K.L., Kenstowicz, M.J. (Eds.), Ken Hale: A Life in Language, Current Studies in Linguistics. MIT Press, Cambridge, Mass, pp. 1–52.
Chomsky, N., 2000. Minimalist Inquiries: The Framework, in: Martin, R., Michaels, D., Uriagereka, J. (Eds.), Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik. MIT Press, Cambridge, Mass, pp. 89–156.
Chomsky, N., 1995a. Language and Nature. Mind, New Series 104, 1–61.
Chomsky, N., 1995b. The Minimalist program, Current studies in linguistics. The MIT Press, Cambridge, Mass.
Chomsky, N., 1995c. The Minimalist Program. MIT Press.
Chomsky, N., 1993a. On the nature, use, and acquisition of language, in: Goldman, A.I. (Ed.), Readings in Philosophy and Cognitive Science. MIT Press, Cambridge, pp. 511–534.
Chomsky, N., 1993b. A minimalist program for linguistic theory, in: Hale, K., Keyser, S.J. (Eds.), The View From Building 20: Essays in Linguistics in Honor of Sylvain Bromberger. MIT Press.
Chomsky, N., 1988. Language and problems of knowledge: the Managua lectures, Current studies in linguistics series. MIT Press, Cambridge, Mass.
Chomsky, N., 1986. Knowledge of language: its nature, origin, and use, Convergence. Praeger, New York.
Chomsky, N., 1982. Some concepts and consequences of the theory of government and binding, Linguistic inquiry monographs. MIT Press, Cambridge, Mass.
Chomsky, N., 1981. Lectures on Government and Binding: The Pisa Lectures, Studies in generative grammar. Foris Publications, Dordrecht, Holland ; Cinnaminson, [N.J.].
Chomsky, N., 1980. Rules and representations, Columbia classics in philosophy. Columbia University Press, New York.
Chomsky, N., 1970a. Current Issues in Linguistic Theory, Minor. Mouton, The Hague.
Chomsky, N., 1970b. Remarks on nominalization, in: Jacobs, R.A., Rosenbaum, P.S. (Eds.), Readings in English Transformational Grammar. Georgetown Univ School of Language, Washington, D.C, pp. 184–221.
Chomsky, N., 1965. Aspects of the theory of syntax, 19. pr. ed, Special technical report of the Research Laboratory of Electronics of the Massachusetts Institute of Technology. MIT Press, Cambridge/Mass.
Chomsky, N., 1957. Syntactic structures, 2nd ed. ed. Mouton de Gruyter, Berlin ; New York.
Church, A., 1936. An Unsolvable Problem of Elementary Number Theory. Journal of Symbolic Logic 1, 73–74.
Churchland, P.M., 2013. Matter and Consciousness, Third edition. ed. MIT Press, Cambridge, Massachusetts.
Churchland, P.M., 2007. Neurophilosophy at Work. Cambridge University Press, Cambridge.
Churchland, P.M., 1993. A Neurocomputational Perspective: The Nature of Mind and the Structure of Science, 3rd pr. ed, A Bradford book. MIT Press, Cambridge, Mass.

Churchland, P.M., 1981. Eliminative Materialism and the Propositional Attitudes. The Journal of Philosophy 78, 67–90. https://doi.org/10.2307/2025900

Churchland, P.S., Sejnowski, T.J., 1999. The Computational Brain, 5. print. ed, Computational neuroscience. MIT Press, Cambridge, Mass.

Cleland, C.E., 2004. The concept of computability. Theoretical Computer Science 317, 209–225. https://doi.org/10.1016/j.tcs.2003.12.012

Cleland, C.E., 2002. On effective procedures. Minds and Machines 12, 159–179.

Cleland, C.E., 2001. Recipes, Algorithms, and Programs. Minds and Machines 11, 219–237. https://doi.org/10.1023/A:1011251504223

Cleland, C.E., 1995. Effective procedures and computable functions. Minds and Machines 5, 9–23.

Cleland, C.E., 1993. Is the Church-Turing Thesis True? Minds and Machines 3, 283–312.

Cohen, J., 1998. Frege and Psychologism. Philosophical Papers 27, 45–67. https://doi.org/10.1080/05568649809506575

Cook, V., Newson, M., 2007. Chomsky's universal grammar: an introduction, 3rd ed. ed. Blackwell Pub, Malden, MA ; Oxford.

Copi, I.M., Cohen, C., McMahon, K.D., 2014. Introduction to logic, 14th ed. Pearson Education Limited.

Cowie, F., 2008. Innateness and Language, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Cowie, F., 1999. What's within? nativism reconsidered, Philosophy of mind series. Oxford University Press, New York.

Cowie, F., 1997. The logical problem of language acquisition. Synthese 111, 17–51.

Crane, T., 2016. The mechanical mind: a philosophical introduction to minds, machines and mental representation, 3rd Edition. ed. Routledge, Taylor & Francis Group, London ; New York.

Crane, T., 1990. The language of thought: No syntax without semantics. Mind and Language 5, 187–213.

Cummins, R., 1983. The Nature of Psychological Explanation. MIT Press, Cambridge, Mass.

Cummins, R., 1977. Programs in the Explanation of Behavior. Philosophy of Science 44, 269–287. https://doi.org/10.1086/288742

Cummins, R., 1975. Functional Analysis. The Journal of Philosophy 72, 741–765. https://doi.org/10.2307/2024640

Cummins, R.C., 2002. Neo-teleology, in: Wright, C., Cummins, R.E., Perlman, M. (Eds.), Functions: New Essays in the Philosophy of Psychology and Biology. Oxford University Press.

Cummins, R.C., Schwarz, G., 1991. Connectionism, computation, and cognition, in: Horgan, T.E., Tienson, J.L. (Eds.), Connectionism and the Philosophy of Mind. Kluwer Academic Publishers, pp. 60--73.

Davidson, D., 1992. The Second Person. Midwest Studies in Philosophy 17, 255–267. https://doi.org/10.1111/j.1475-4975.1992.tb00154.x

Davidson, D., 1970. Mental Events, in: Foster, L., Swanson, J.W. (Eds.), Essays on Actions and Events. Clarendon Press, pp. 207–224.

Davis, M., 1982. Computability & unsolvability, Dover ed. ed. Dover, New York.

Davis, M., Sigal, R., Weyuker, E.J., 1994. Computability, complexity, and languages: fundamentals of theoretical computer science, 2nd. ed. ed, Computer science and scientific computing. Academic Press, Harcourt, Brace, Boston.

Demey, L., Smessaert, H., 2016. The Interaction Between Logic and Geometry in Aristotelian Diagrams, in: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (Eds.), Diagrammatic Representation and Inference. Springer International Publishing, pp. 67–82.

Dennett, D.C., 2006. Sweet dreams: philosophical obstacles to a science of consciousness. MIT Press, London.
Dennett, D.C., 1996. Darwin's dangerous idea: evolution and the meanings of life, A Touchstone book. Touchstone, New York.
Dennett, D.C., 1991. Consciousness explained, 1st ed. ed. Little, Brown and Co, Boston.
Dennett, D.C., 1987. The Intentional Stance. MIT Press.
Dillard, P., 2014. Ross Revisited. American Catholic Philosophical Quarterly 88, 139–147.
Dillard, P., 2011. Two Unsuccessful Arguments for Immaterialism. American Catholic Philosophical Quarterly 85, 269–286.
Dretske, F., 1993. The Nature of Thought. Philosophical Studies 70, 185–99.
Dretske, F., 1988. Explaining Behavior: Reasons in a World of Causes. MIT Press.
Dretske, F., 1986. Misrepresentation, in: Bogdan, R. (Ed.), Belief: Form, Content, and Function. Oxford University Press, pp. 17--36.
Dretske, F., 1981. Knowledge and the Flow of Information. MIT Press.
Dreyfus, H.L., 1992. What Computers Still Can't Do: A Critique of Artificial Reason. MIT Press, Cambridge, Mass.
Dreyfus, H.L., 1979. What Computers Can't Do: The Limits of Artificial Intelligence, Rev. ed. ed, Harper Colophon books. Harper & Row, New York.
Ebbs, G., 2001. Rule-following and Realism. Harvard University Press, Cambridge, Mass.
Eliasmith, C., 2003. Moving beyond Metaphors: Understanding the Mind for What It Is. The Journal of Philosophy 100, 493–520.
Evans, G., 1975. Identity and Predication. The Journal of Philosophy 72, 343–363. https://doi.org/10.2307/2025212
Feser, E., 2017. Revisiting Ross on the immateriality of thought. Edward Feser's Blog.
Feser, E., 2016. Putnam and analytical Thomism, Part I. Edward Feser's Blog.
Feser, E., 2014. Scholastic Metaphysics: A Contemporary Introduction, Editiones scholasticae. Ed. Scholasticae, Heusenstamm.
Feser, E., 2013. Kripke, Ross, and the Immaterial Aspects of Thought. American Catholic Philosophical Quarterly 87, 1–32.
Feser, E., 2009. Aquinas: A Beginner's Guide, Oneworld beginner's guides. Oneworld, Oxford.
Feser, E., 2006. Philosophy of Mind: A Beginner's Guide, Rev. ed. ed. Oneworld, London.
Feser, E., n.d. From Aristotle to John Searle and Back Again: Formal Causes, Teleology, and Computation in Nature. Formal Causes 36.
Field, H., 1998. Mathematical Objectivity and Mathematical Objects, in: C. MacDonald, S.L. (Ed.), Contemporary Readings in the Foundations of Metaphysics. Blackwell, pp. 387–403.
Field, H.H., 1989. Realism, Mathematics, and Modality. Blackwell, Oxford, UK ; New York, NY, USA.
Field, H.H., 1980. Mathematics Without Numbers: A Defence of Nominalism. Blackwell, Oxford.
Fodor, J.A., 2008. LOT 2: The language of thought revisited. Clarendon Press ; Oxford University Press, Oxford : New York.
Fodor, J.A., 2000. The mind doesn't work that way: the scope and limits of computational psychology, Representation and mind. MIT Press, Cambridge, Mass.
Fodor, J.A., 1998. Concepts: where cognitive science went wrong, Oxford cognitive science series. Clarendon Press ; Oxford University Press, Oxford : New York.
Fodor, J.A., 1996. Fodor, Jerry A., in: Guttenplan, S.D. (Ed.), A Companion to the Philosophy of Mind, Blackwell Companions to Philosophy. Blackwell, Oxford, pp. 292–300.
Fodor, J.A., 1994. A theory of content: and other essays, Reprint. ed. Press, Cambridge, Mass.

Fodor, J.A., 1989. Psychosemantics: the problem of meaning in the philosophy of mind, First MIT Press paperback edition. ed, Explorations in cognitive science. MIT Press, Cambridge, Massachusetts London.

Fodor, J.A., 1985. Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum. Mind 94, 76–100.

Fodor, J.A., 1983. The Modularity of Mind: An essay on Faculty Psychology. MIT Press, Cambridge, Mass.

Fodor, J.A., 1981. The mind-body problem. Sci. Am. 244, 114–120, 122–123.

Fodor, J.A., 1975. The Language of Thought. Harvard University Press.

Fodor, J.A., 1968a. Psychological Explanation: An Introduction To The Philosophy Of Psychology. Random House.

Fodor, J.A., 1968b. The appeal to tacit knowledge in psychological explanation. Journal of Philosophy 65, 627–40.

Fodor, J.A., 1965. Explanations in psychology, in: Black, M. (Ed.), Philosophy in America. Cornell University Press, pp. 161--179.

Fodor, J.A., Pylyshyn, Z.W., 1988. Connectionism and cognitive architecture. Cognition 28, 3–71.

Foellesdal, D., 1973. Indeterminacy of Translation and Under-Determination of the Theory of Nature. Dialectica 27, 289–301.

Franklin, J., 2014. An Aristotelian Realist Philosophy of Mathematics. Palgrave Macmillan UK, London. https://doi.org/10.1057/9781137400734

Frege, G., 1980. The foundations of arithmetic: a logico-mathematical enquiry into the concept of number, 2nd rev. ed. ed. Northwestern University Press, Evanston, Ill.

Garson, J., 2018. Connectionism, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Garvey, J. (Ed.), 2011. The Continuum companion to philosophy of mind, Continuum companions. Continuum, London ; New York.

Gennaro, R.J. (Ed.), 2018. The Routledge handbook of consciousness, 1 [edition]. ed, Routledge handbooks in philosophy. Routledge, New York.

George, A., 1985. Skolem and the löwenheim-skolem theorem: a case study of the philosophical significance of mathematical results. History and Philosophy of Logic 6, 75–89.

Gillett, C., 2007. A Mechanist Manifesto for the Philosophy of Mind: A Third Way for Functionalists. Journal of Philosophical Research 32, 21–42. https://doi.org/10.5840/jpr20073245

Gluer, K., Wikforss, A., 2009. Against Content Normativity. Mind 118, 31–70. https://doi.org/10.1093/mind/fzn154

Göcke, B.P. (Ed.), 2012. After physicalism. University of Notre Dame Press, Notre Dame, Ind.

Gödel, K., 1934. On Undecidable Propositions of Formal Mathematical Systems, in: Davis, M. (Ed.), Reprinted in: The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems, and Computable Functions. 2004., Dover Books on Mathematics. Dover Publication, Mineola, NY.

Gödel, K., 1931. On Formally Undecidable Propositions of Principia Mathematica and Related Systems I. Monatshefte für Mathematik und Physik 38–38, 173–198. https://doi.org/10.1007/BF01700692

Godfrey-Smith, P., 2009. Triviality arguments against functionalism. Philosophical Studies 145, 273–295. https://doi.org/10.1007/s11098-008-9231-3

Goodman, N., 1983. Fact, fiction, and forecast, 4th ed. ed. Harvard University Press, Cambridge, Mass.

Hacker, P.M.S., 2017. An Intellectual Entertainment: Thought and Language. Philosophy 92, 271–296. https://doi.org/10.1017/S0031819116000450

Hacker, P.M.S., 2013. The intellectual powers: a study of human nature. Wiley Blackwell, Chichester, West Sussex ; Malden, MA.

Hacker, P.M.S., 2007. Human nature: the categorial framework. Blackwell Pub, Malden, MA ; Oxford.

Hacker, P.M.S., 1990. Chomsky's problems. Language & Communication 10, 127–148. https://doi.org/10.1016/0271-5309(90)90030-F

Hacker, P.M.S., Baker, G.P., 2014. Wittgenstein: rules, grammar and necessity: essays and exegesis of ″ 185-242. Wiley-Blackwell, Chichester.

Hacker, P.M.S., Baker, G.P., 2005. Wittgenstein: understanding and meaning, 2nd, extensively rev. ed ed, Analytical commentary on the Philosophical investigations. Blackwell Pub, Oxford, UK ; Malden, MA.

Hacker, P.M.S., Baker, G.P., 1984. Scepticism, rules and language. B. Blackwell, Oxford, Oxfordshire.

Hacking, I., 2014. Why is There Philosophy of Mathematics at All? Cambridge University Press.

Hacking, I., 1993. On Kripke's and Goodman's uses of "grue." Philosophy 68, 269–295.

Hacking, I., 1979. What is logic? Journal of Philosophy 76, 285–319.

Harnad, S., 2012a. The Mind/Body Problem is The Feeling/Function Problem: Harnad on Dennett on Chalmers. [Unpiblished].

Harnad, S., 2012b. Alan Turing and the "Hard" and "Easy" Problem of Cognition: Doing and Feeling. CoRR abs/1206.3658.

Harte, V., 2002. Plato on parts and wholes: the metaphysics of structure. Clarendon Press ; Oxford University Press, Oxford : New York.

Haugeland, J. (Ed.), 1997. Mind design II: philosophy, psychology, artificial intelligence, Rev. and enl. ed. ed. MIT Press, Cambridge, Mass.

Haugeland, J., 1981. Semantic engines: An introduction to mind design, in: Haugeland, J. (Ed.), Mind Design I: Philosophy, Psychology, Artificial Intelligence. MIT Press, pp. 1–34.

Haugeland, J., 1978. The nature and plausibility of Cognitivism. Behavioral and Brain Sciences 1, 215–226. https://doi.org/10.1017/S0140525X00074148

Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? Science 298, 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Heil, J., 2013. Philosophy of mind: a contemporary introduction, 3. ed. ed, Routledge contemporary introductions to philosophy. Routledge, New York, NY.

Heil, J. (Ed.), 2004. Philosophy of mind: a guide and anthology. Oxford University Press, Oxford ; New York.

Hodges, W., Scanlon, T., 2018. First-order Model Theory, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Hoffman, P., 1985. Kripke on private language. Philosophical Studies 47, 23–28. https://doi.org/10.1007/BF00355085

Hornstein, N., Nunes, J., Grohmann, K.K., 2005. Understanding minimalism, Cambridge textbooks in linguistics. Cambridge University Press, Cambridge, UK ; New York.

Horst, S., Zalta, E.N., 2003. The Computational Theory of Mind, in: The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Horsten, L., 2018. Philosophy of Mathematics, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Horsten, L., 1995. The church-Turing thesis and effective mundane procedures. Minds and Machines 5, 1–8.
Horsten, L., Welch, P. (Eds.), 2016. Gödel's disjunction: the scope and limits of mathematical knowledge, First edition. ed. Oxford University Press, Oxford.
Horwich, P., 1995. Meaning, Use and Truth: On Whether a Use-Theory of Meaning Is Precluded by the Requirement that Whatever Constitutes the Meaning of a Predicate Be Capable of Determining the Set of Things of Which the Predicate is True and to Which It Ought to be Applied. Mind 104, 355–368.
Hunston, S., Francis, G., 2000. Pattern grammar: a corpus-driven approach to the lexical grammar of English, Studies in corpus linguistics. John Benjamins Pub. Co, Amsterdam ; Philadelphia.
Husserl, E., 1982. Logical investigations. Routledge & Kegan Paul, London.
Hylton, P., 2018. Willard van Orman Quine, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.
Israel, D., 2002. Reflections on Gödel's and Gandy's Reflections on Turing's Thesis. Minds and Machines 12, 181–201. https://doi.org/10.1023/A:1015634729532
Jackendoff, R., 1977. X syntax: a study of phrase structure, Linguistic inquiry monographs. MIT Press, Cambridge, Mass.
Jackson, F. (Ed.), 1998. Consciousness, The international research library of philosophy. Ashgate : Dartmouth; Ashgate, Aldershot, England : Brookfield, Vt.
Jackson, F., Smith, M. (Eds.), 2007. The Oxford Handbook of Contemporary Philosophy, Reprint. ed. Oxford University Press, Oxford.
Jaworski, W., 2016. Structure and the metaphysics of mind: how hylomorphism solves the mind-body problem, First edition. ed. Oxford University Press, Oxford.
Jaworski, W., 2011. Philosophy of mind: a comprehensive introduction. Wiley-Blackwell, Chichester, West Sussex ; Malden, MA.
Kasher, A., Chomsky, N. (Eds.), 1992. The Chomskyan turn, Reprinted. ed. Blackwell, Oxford.
Kenny, A., 1971. The Homunculus Fallacy, in: Grene, M.G., Prigogine, I. (Eds.), Interpretations of Life and Mind. Humanities Press, pp. 155–165.
Kim, J., 2011. Philosophy of mind, 3rd ed. ed. Westview Press, Boulder, CO.
Kim, J., 2008. Physicalism, or something near enough. Princeton University Press, Princeton, N.J.; Woodstock.
Kim, J., 1998. Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation, 1. paperback ed. ed, Representation and mind. MIT Press, Cambridge, Mass.
Kim, J., 1993. Supervenience and Mind: Selected Philosophical Essays, Cambridge Studies in Philosophy. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511625220
Kirk, R., 2015. Zombies, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.
Koellner, P., 2018a. On the Question of Whether the Mind Can Be Mechanized, I: From Gödel to Penrose: The Journal of Philosophy 115, 337–360. https://doi.org/10.5840/jphil2018115721
Koellner, P., 2018b. On the Question of Whether the Mind Can Be Mechanized, II: Penrose's New Argument: The Journal of Philosophy 115, 453–484. https://doi.org/10.5840/jphil2018115926
Koslicki, K., 2008. The structure of objects. Oxford University Press, Oxford ; New York.
Kripke, S.A., 1984. Lessons on Functionalism and Automata.

Kripke, S.A., 1982. Wittgenstein on rules and private language: an elementary exposition, Repr. ed. Harvard Univ. Press, Cambridge, Mass.

Kripke, S.A., 1981. Wittgenstein on Rules and Private Language, in: Block, I. (Ed.), Perspectives on the Philosophy of Wittgenstein. MIT Press, Cambridge, Mass, pp. 238–312.

Kurzweil, R., 2012. How to create a mind: the secret of human thought revealed. Viking, New York.

Kusch, M., 2015. Psychologism, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Kusch, M., 2010. Rule-Following. https://doi.org/10.1093/obo/9780195396577-0106

Kusch, M., 2007. Rule Skepticism: Searle's criticism of Kripke's Wittgenstein, in: Tsohatzidis, S.L. (Ed.), John Searle's Philosophy of Language: Force, Meaning, and Mind. Cambridge University Press, Cambridge ; New York.

Kusch, M., 2006. A sceptical guide to meaning and rules: defending Kripke's Wittgenstein. McGill-Queen's University Press, Montreal ; Ithaca, NY.

Ladyman, J., 2009. What does it mean to say that a physical system implements a computation? Theoretical Computer Science 410, 376–383. https://doi.org/10.1016/j.tcs.2008.09.047

Lande, K.J., 2011. Reazlining Functions? An Argument against Functionalism. http://www.bu.edu/conscious/2011papers/Lande.pdf.

Leuenberger, S., 2008. Supervenience in Metaphysics. Philosophy Compass 3, 749–762. https://doi.org/10.1111/j.1747-9991.2008.00150.x

Levin, J., 2018. Functionalism, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Levine, J., 2001. Purple haze: the puzzle of consciousness. Oxford University Press, Oxford [England]; New York.

Lewis, D., 1979. Lucas against Mechanism II. Canadian Journal of Philosophy 9, 373–376.

Lewis, D., 1972. Psychophysical and theoretical identifications. Australasian Journal of Philosophy 50, 249–258.

Lewis, D., 1969. Lucas against Mechanism. Philosophy 44, 231–233.

Lonergan, B.J.F., 1992. Insight: A Study of Human Understanding, Collected works of Bernard Lonergan. Published by University of Toronto Press for Lonergan Research Institute of Regis College, Toronto.

Lucas, J.R., 1996. Minds, Machines, and Gödel: A Retrospect, in: Millican, P.J.R., Clark, A. (Eds.), Etica E Politica. Clarendon Press, p. 1.

Lucas, J.R., 1984. Lucas against Mechanism II: A Rejoinder. Canadian Journal of Philosophy 14, 189–191.

Lucas, J.R., 1961. Minds, Machines and Gödel. Philosophy 36, 112–127.

Ludlow, P., Nagasawa, Y., Stoljar, D. (Eds.), 2004. There's something about Mary: essays on phenomenal consciousness and Frank Jackson's knowledge argument. MIT Press, Cambridge, Mass.

Macdonald, G., Papineau, D. (Eds.), 2006. Teleosemantics: New Philosophical Essays. Clarendon Press, Oxford ; New York.

MacIntyre, A., 1986. The intelligibility of action, in: Margolis, J., Krausz, M., Burian, R.M. (Eds.), Rationality, Relativism, and the Human Sciences. M. Nijhoff, pp. 63--80.

MacIntyre, A.C., 2006. Selected essays. Cambridge University Press, Cambridge, UK ; New York.

Maley, C., Piccinini, G., 2013. Get the Latest Upgrade: Functionalism 6.3.1. Philosophia Scientae 135–149. https://doi.org/10.4000/philosophiascientiae.861

Marcus, R., 2017. The Indispensability Argument in the Philosophy of Mathematics. The Internet Encyclopedia of Philosophy.

Marr, D., 1982. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, 2010 edition. ed. MIT Press, Cambridge, Mass.
Mcdowell, J., 1992. Meaning and Intentionality in Wittgenstein's Later Philosophy. Midwest Studies in Philosophy 17, 40–52. https://doi.org/10.1111/j.1475-4975.1992.tb00141.x
McDowell, J., 1984. Wittgenstein on following a rule. Synthese 58, 325–364.
McGilvray, J.A. (Ed.), 2005. The Cambridge companion to Chomsky. Cambridge University Press, Cambridge.
McLaughlin, B.P., 2006. Is Role-Functionalism Committed to Epiphenomenalism? Journal of Consciousness Studies 13, 39–66.
McLaughlin, B.P., Beckermann, A., Walter, S. (Eds.), 2009. The Oxford handbook of philosophy of mind, Oxford handbooks in philosophy. Claredon Press ; Oxford University Press, Oxford : New York.
McLaughlin, B.P., Cohen, J.D. (Eds.), 2007. Contemporary debates in philosophy of mind, Contemporary debates in philosophy. Blackwell Pub, Malden, Mass.
Mendelson, E., 2015. Introduction to mathematical logic, Sixth edition. ed, Textbooks in mathematics. CRC Press/Taylor & Francis Group, Boca Raton.
Miłkowski, M., 2013. Explaining the computational mind. The MIT Press, Cambrid, Massachusetts.
Milkowski, M., Talmont-Kaminski, K., 2010. Beyond Description: Naturalism and Normativity. College Publications, London.
Miller, A., 2007. Philosophy of language. Routledge, Milton Park, Abingdon, Oxon; New York.
Miller, A., Wright, C. (Eds.), 2002. Rule-following and meaning. Acumen, Chesham.
Millikan, R.G., 1990. Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox. The Philosophical Review 99, 323–353. https://doi.org/10.2307/2185347
Minsky, M., 1967. Computation: finite and infinite machines. Prentice-Hall, Englewood Cliffs, NJ.
Mohanty, J.N., 1997. The Concept of "Psychologism" in Frege and Husserl. Philosophy & Rhetoric 30, 271–290.
Morris, M., 2006. An Introduction to the Philosophy of Language. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511801464
Moser, P.K., Trout, J.D. (Eds.), 1995. Contemporary Materialism: A Reader. Routledge, London ; New York.
Nagel, T., 2012. Mind and cosmos: why the materialist neo-Darwinian conception of nature is almost certainly false. Oxford University Press, New York.
Nagel, T., 1986. The View From Nowhere. Oxford University Press.
Neander, K., 2018. Teleological Theories of Mental Content, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.
Nicholas Rescher, Paul Oppenheim, 1955. Logical Analysis of Gestalt Concepts. The British Journal for the Philosophy of Science 6.
Nida-Rümelin, M., 2015. Qualia: The Knowledge Argument, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.
Niiniluoto, I., 2002. Critical scientific realism, Clarendon library of logic and philosophy. Oxford University Press, Oxford.
Notturno, M.A. (Ed.), 1989. Perspectives on psychologism, Brill's studies in epistemology, psychology, and psychiatry. E.J. Brill, Leiden ; New York.
Nozick, R., 2001. Invariances: the structure of the objective world. Belknap Press of Harvard University Press, Cambridge, Mass.
O'Connor, T., Robb, D., Heil, J. (Eds.), 2003. Philosophy of mind: contemporary readings, Routledge contemporary readings in philosophy. Routledge, London ; New York.

Oderberg, D.S., 2014. Is Form Structure?, in: Novotny, D.D., Novak, L. (Eds.), Neo-Aristotelian Perspectives in Metaphysics. Routledge, London, pp. 164–180.

Oderberg, D.S., 2007. Real Essentialism. Routledge.

Papineau, D., 2016. Naturalism, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Papineau, D., 1993. Philosophical naturalism. B. Blackwell, Oxford, UK ; Cambridge, Mass., USA.

Paseau, A., 2016. Naturalism in the Philosophy of Mathematics, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Pasnau, R., 1998. Aquinas and the Content Fallacy. Modern Schoolman 75, 293–314.

Pelletier, F.J., Elio, R., Hanson, P., 2008. Is Logic all in our Heads? From Naturalism to Psychologism. Studia Logica 88, 3–66.

Penrose, R., 2011a. Godel, the Mind, and the Laws of Physics, in: Baaz, M., Putnam, H., Papadimitriou, C.H., Scott, D.S., Harper, C.L. (Eds.), Kurt Gödel and the Foundations of Mathematics: Horizons of Truth. Cambridge University Press, New York, pp. 339–358.

Penrose, R., 2011b. Godel, the Mind, and the Laws of Physics, in: Baaz, M. (Ed.), Kurt Gödel and the Foundations of Mathematics: Horizons of Truth. Cambridge University Press, p. 339.

Penrose, R., 2002. Consciousness, Computation, and the Chinese Room, in: Preston, J., Bishop, M. (Eds.), Views into the Chinese Room: New Essays on Searle and Artificial Intelligence. Clarendon Press, Oxford ; New York.

Penrose, R., 1999. The emperor's new mind concerning computers, minds, and the laws of physics. Oxford Univ. Press, Oxford.

Penrose, R., 1996a. Shadows of the mind: a search for the missing science of consciousness, 1. paperback ed. ed. Oxford Univ. Press, Oxford.

Penrose, R., 1996b. Beyond the Doubting of a Shadow: A Reply to Commentaries on Shadows of the Mind. Psyche 23.

Penrose, R., 1996c. Mathematical Intelligence, in: Khalfa, J. (Ed.), What Is Intelligence?, The Darwin College Lectures. Cambridge Univ. Press, Cambridge, pp. 107–137.

Piccinini, G., 2017. Computation in Physical Systems, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Piccinini, G., 2015. The nature of computation. Oxford University Press, New York, NY.

Piccinini, G., 2010a. The Resilience of Computationalism. Philosophy of Science 77, 852–861. https://doi.org/10.1086/656549

Piccinini, G., 2010b. The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism. Philosophy and Phenomenological Research 81, 269–311. https://doi.org/10.1111/j.1933-1592.2010.00356.x

Piccinini, G., 2009. Computationalism in the Philosophy of Mind. Philosophy Compass 4, 515–532. https://doi.org/10.1111/j.1747-9991.2009.00215.x

Piccinini, G., 2008a. Some neural networks compute, others don't. Neural Networks 21, 311–321. https://doi.org/10.1016/j.neunet.2007.12.010

Piccinini, G., 2008b. Computation without Representation. Philosophical Studies 137, 205–241. https://doi.org/10.1007/s11098-005-5385-4

Piccinini, G., 2004a. Functionalism, Computationalism, and Mental Contents. Canadian Journal of Philosophy 34, 375–410. https://doi.org/10.1080/00455091.2004.10716572

Piccinini, G., 2004b. Functionalism, computationalism, and mental states. Studies in History and Philosophy of Science Part A 35, 811–833. https://doi.org/10.1016/j.shpsa.2004.02.003

Piccinini, G., Scarantino, A., 2011. Information processing, computation, and cognition. Journal of Biological Physics 37, 1–38.

Pinker, S., 2015. How the mind works. Penguin Books, London.
Pinker, S., 2003. The Blank Slate: The Modern Denial of Human Nature, Reprint edition. ed. Penguin Books.
Plantinga, A., 2011. Where the Conflict Really Lies: Science, Religion, and Naturalism. Oxford University Press, New York.
Polger, T.W., 2015. Realization and Multiple Realization, Chicken and Egg: Realization and Multiple Realization, Chicken and Egg. European Journal of Philosophy 23, 862–877. https://doi.org/10.1111/ejop.12017
Preston, J., Bishop, M. (Eds.), 2002. Views into the Chinese room: new essays on Searle and artificial intelligence. Clarendon Press, Oxford ; New York.
Putnam, H., 2002. Comments on John Haldane's Paper, in: Żegleń, U.M., Conant, J. (Eds.), Hilary Putnam: Pragmatism and Realism. Routledge, London; New York, pp. 105–108.
Putnam, H., 1988. Representation and reality, Representation and mind. MIT Press, Cambridge, Mass.
Putnam, H., 1967a. The mental life of some machines, in: Castaneda, H.-N. (Ed.), Intentionality, Minds and Perception. Wayne State University Press.
Putnam, H., 1967b. Psychological predicates, in: Capitan, W.H., Merrill, D.D. (Eds.), Art, Mind, and Religion. University of Pittsburgh Press, pp. 37--48.
Putnam, H., 1960. Minds and machines, in: Hook, S. (Ed.), Journal of Symbolic Logic. New York University Press, pp. 57–80.
Quine, W.V., 1976. The Ways of Paradox and Other Essays, Rev. and enl. ed., 7. printing. ed. Harvard University Press, Cambridge, Mass.
Quine, W.V., 1970. On the Reasons for Indeterminacy of Translation. The Journal of Philosophy 67, 178–183. https://doi.org/10.2307/2023887
Quine, W.V., 1969. Ontological relativity and other essays, 10th-11. print ed. Columbia University Press, New York.
Quine, W.V., 1960. Word and object, New ed. ed. MIT Press, Cambridge, Mass.
Reppert, V., 2009. The Argument from Reason, in: Craig, W.L., Moreland, J.P. (Eds.), The Blackwell Companion to Natural Theology. Wiley-Blackwell, Chichester, U.K. ; Malden, MA, pp. 344–390.
Rescorla, M., 2017. The Computational Theory of Mind, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.
Rescorla, M., 2015. The Representational Foundations of Computation. Philosophia Mathematica 23, 338–366.
Rescorla, M., 2014. The Causal Relevance of Content to Computation. Philosophy and Phenomenological Research 88, 173–208.
Rescorla, M., 2013. Against Structuralist Theories of Computational Implementation. British Journal for the Philosophy of Science 64, 681–707.
Rescorla, M., 2012. Are Computational Transitions Sensitive to Semantics? Australasian Journal of Philosophy 90, 703–721.
Rey, G., 1997. Contemporary philosophy of mind: a contentiously classical approach, Contemporary philosophy. Blackwell, Cambridge, Mass.
Robinson, H. (Ed.), 1996. Objections to Physicalism. Clarendon, Oxford.
Roeper, T., Williams, E., 1987. Parameter Setting. Springer Netherlands, Dordrecht.
Rosen, G., 2018. Abstract Objects, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Ross, J., 2012. Merely Metaphysical Possibility, in: Doolan, G.T. (Ed.), The Science of Being as Being, Metaphysical Investigations. Catholic University of America Press, pp. 154–174. https://doi.org/10.2307/j.ctt28504n.10

Ross, J., 1992. Immaterial Aspects of Thought. The Journal of Philosophy 89, 136. https://doi.org/10.2307/2026790

Ross, J.F., 2008. Thought and world: the hidden necessities. University of Notre Dame Press, Notre Dame, Ind.

Ross, J.F., 1993. Musical Standards as a Function of Musical Accomplishment, in: Krausz, M. (Ed.), The Interpretation of Music: Philosophical Essays. Clarendon Press ; Oxford University Press, Oxford : Oxford ; New York, pp. 89–102.

Ross, J.F., 1990. The Fate of the Analysis: Aristotle's Revenge. Proceedings of the American Catholic Philosophical Association 64, 51–74. https://doi.org/10.5840/acpaproc19906424

Ross, J.F., 1989. The Crash of Modal Metaphysics. The Review of Metaphysics 43, 251–279.

Ross, J.F., 1988. Eschatological Pragmatism, in: Morris, T.V. (Ed.), Philosophy and the Christian Faith. Univ. Of Notre Dame Press, pp. 279--300.

Ross, J.F., 1986. God, Creator of Kinds and Possibilities, in: Wainwright, W., Audi, R. (Eds.), Rationality, Religious Belief, and Moral Commitment: New Essays in the Philosophy of Religion. Cornell University Press, pp. 315--334.

Searle, J.R., 2002. Consciousness and Language. Cambridge University Press, New York.

Searle, J.R., 1992. The rediscovery of the mind, Representation and mind. MIT Press, Cambridge, Mass.

Searle, J.R., 1990. Is the Brain a Digital Computer? Proceedings and Addresses of the American Philosophical Association 64, 21–37. https://doi.org/10.2307/3130074

Searle, J.R., 1987. Indeterminacy, Empiricism, and the First Person. The Journal of Philosophy 84, 123. https://doi.org/10.2307/2026595

Searle, J.R., 1983. Intentionality, an essay in the philosophy of mind. Cambridge University Press, Cambridge [Cambridgeshire] ; New York.

Searle, J.R., 1980. Minds, brains, and programs. Behavioral and Brain Sciences 3, 417. https://doi.org/10.1017/S0140525X00005756

Searle, J.R., 1970. Speech Acts: An Essay in the Philosophy of Language, n Reprint edition. ed. Cambridge University Press, Cambridge.

Seligman, J., 2002. The Scope of Turing's Analysis of Effective Procedures. Minds and Machines 12, 203–220. https://doi.org/10.1023/A:1015638814511

Shanks, N., Brzeziński, J. (Eds.), 1998. Idealization in contemporary physics, Idealization. Rodopi, Amsterdam.

Shapiro, L.A., 2008. How to Test for Multiple Realization. Philosophy of Science 75, 514–525. https://doi.org/10.1086/594503

Shapiro, L.A., 2000. Multiple realizations. Journal of Philosophy 97, 635–654.

Shapiro, S., 2018. Mathematical Structuralism [WWW Document]. Internet Encyclopedia of Philosophy: A Peer Reviewed Academic Resource. URL https://www.iep.utm.edu/m-struct/ (accessed 11.15.18).

Shapiro, S. (Ed.), 2007. The Oxford handbook of philosophy of mathematics and logic, Oxford handbooks in philosophy. Oxford University Press, Oxford ; New York.

Shapiro, S., 2001. Computer Sience: The Study of Procedures. https://cse.buffalo.edu/~shapiro/Papers/whatiscs.pdf.

Shoemaker, S., 2007. Physical realization. Oxford University Press, Oxford ; New York.

Smart, J.J.C., 1959. Sensations and Brain Processes. The Philosophical Review 68, 141–156. https://doi.org/10.2307/2182164

Smith, P., 2013. An introduction to Gödel's theorems, 2nd ed. ed, Cambridge introductions to philosophy. Cambridge University Press, Cambridge, England ; New York.

Soames, S., 1999. The Indeterminacy of Translation and the Inscrutability of Reference. Canadian Journal of Philosophy 29, 321–370.

Soames, S., 1998a. Skepticism about meaning, indeterminacy, normativity, and the rule-following paradox. Canadian Journal of Philosophy, Supp 23, 211--50.

Soames, S., 1998b. Facts, Truth Conditions, and the Skeptical Solution to the Rule-Following Paradox. Philosophical Perspectives 12, 313–348.

Sprevak, M., 2010. Computation, individuation, and the received view on representation. Studies in History and Philosophy of Science Part A 41, 260–270.

Stabler, E.P., 1987. Kripke on functionalism and automata. Synthese 70, 1–22.

Stabler, E.P., 1986. Computational models of language processing. Behavioral and Brain Sciences 9, 550.

Stewart, I., 2001. Easter Is a Quasicrystal. Scientific American 284, 80–83.

Stokhof, M., Lambalgen van, M., 2017. What Cost Naturalism?, in: Balogh, K., Petersen, W. (Eds.), Bridging Formal and Conceptual Semantics: Selected Papers of BRIDGE-14, Studies in Language and Cognition. DUP, Düsseldorf University Press, Düsseldorf, pp. 91–119.

Stokhof, M., van Lambalgen, M., 2011a. Comments-to-comments. Theoretical Linguistics 37, 79–94. https://doi.org/10.1515/thli.2011.008

Stokhof, M., van Lambalgen, M., 2011b. Abstractions and idealisations: The construction of modern linguistics. Theoretical Linguistics 37, 1–26. https://doi.org/10.1515/thli.2011.001

Strawson, G., 2008. Real Materialism and Other Essays. Clarendon Press ; Oxford University Press, Oxford : New York.

Talmont-Kaminski, K., Miłkowski, M. (Eds.), 2013. Regarding the Mind, Naturally: Naturalist Approaches to the Sciences of the Mental. Cambridge Scholars Publishing, Newcastle upon Tyne, UK.

Turing, A., 1936. On Computable Numbers, with an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society 42, 230–265.

Van Gulick, R., 2009. Functionalism, in: McLaughlin, B., Beckermann, A., Walter, S. (Eds.), The Oxford Handbook of Philosophy of Mind. Oxford University Press.

Velmans, M., Schneider, S. (Eds.), 2007. The Blackwell Companion to Consciousness. Wiley, Chichester.

Wikforss, Å., 2018. Does Semantics Need Normativity? Comments on Allan Gibbard: Meaning and Normativity. Inquiry 61, 755–766. https://doi.org/10.1080/0020174X.2018.1424528

William, P.T., 1994. On proof and progress in mathematics.

Wittgenstein, L., 2009. Philosophical Investigations, Rev. 4th ed. ed. Wiley-Blackwell, Chichester, West Sussex, U.K. ; Malden, MA.

Wittgenstein, L., 1996. Remarks on the foundations of mathematics, Rev. ed., 4. print. ed. MIT Press, Cambridge, Mass.

Wolfram, S., 2002. A New Kind of Science. Wolfram Media, Champaign, IL.

Wright, C., 2007. Rule-following without reasons: Wittgenstein's quietism and the constitutive question. Ratio 20, 481–502.

Wright, C., 1981. Rule-following, objectivity and the theory of meaning, in: Holtzman, S.H., Leich, C.M. (Eds.), Wittgenstein: To Follow A Rule. Routledge.

Wright, C., Wright, C., 2017. Indeterminacy of Translation, in: Hale, B., Miller, A. (Eds.), A Companion to the Philosophy of Language. John Wiley & Sons, Ltd, Chichester, UK, pp. 670–702. https://doi.org/10.1002/9781118972090.ch26

Wright, C.J.G., 1989. Wittgenstein's rule-following considerations and the central project of theoretical linguistics, in: George, A. (Ed.), Reflections on Chomsky. Blackwell, pp. 233–264.

Zelazo, P.D., Moscovitch, M., Thompson, E. (Eds.), 2007. The Cambridge handbook of consciousness. Cambridge University Press, Cambridge ; New York.